

NOTES ON GALERKIN-FINITE ELEMENT METHODS FOR THE SHALLOW WATER EQUATIONS WITH CHARACTERISTIC BOUNDARY CONDITIONS.

D.C. ANTONOPOULOS AND V.A. DOUGALIS

ABSTRACT. We consider the shallow water equations in the supercritical and subcritical cases in one space variable, posed in a finite spatial interval with characteristic boundary conditions at the endpoints, which, as is well known, are transparent, i.e. allow outgoing waves to exit without generating spurious reflected waves. Assuming that the resulting initial-boundary-value problems have smooth solutions, we approximate them in space using standard Galerkin-finite element methods and prove L^2 error estimates for the semidiscrete problems on quasiuniform meshes. We discretize the problems in the temporal variable using an explicit, fourth-order accurate Runge-Kutta scheme and check, by means of numerical experiment, that the resulting fully discrete schemes have excellent absorption properties.

1. INTRODUCTION

In this paper we consider the system of *shallow water equations*

$$\begin{aligned}\eta_t + u_x + (\eta u)_x &= 0, \\ u_t + \eta_x + uu_x &= 0,\end{aligned}\tag{1.1}$$

a well known approximation of the two-dimensional Euler equations of water-wave theory, modelling two-way propagation of long surface waves of finite amplitude in a uniform horizontal channel of finite depth, [Wh]. The variables in (1.1) are non-dimensional and unscaled; $x \in \mathbb{R}$ and $t \geq 0$ are proportional to position along the channel and time, respectively, while $\eta = \eta(x, t)$ and $u = u(x, t)$ are proportional to the elevation of the free surface above a level of rest and to the horizontal velocity of the fluid, respectively. The latter is depth-independent to the order of approximation represented by the scaled analog of (1.1). In the variables of (1.1) the bottom of the channel lies at a depth equal to -1 .

It is well known that the initial-value problem for (1.1), posed with smooth initial data $\eta(x, 0)$ and $u(x, 0)$ for $x \in \mathbb{R}$, has, in general, smooth solutions only locally in t , cf. e.g. [Ma], Ch. 2. In this paper we will pose (1.1) in the finite ‘channel’ $[0, L]$ with given initial values at $t = 0$,

$$\eta(x, 0) = \eta^0(x), \quad u(x, 0) = u^0(x), \quad 0 \leq x \leq L,\tag{1.2}$$

and consider *transparent boundary conditions* at $x = 0$ and $x = L$, i.e. conditions that permit the waves to exit the ‘computational’ domain $[0, L]$ without generating spurious reflected waves that pollute the solution inside $[0, L]$. The transparent boundary conditions that we will use are *nonlinear characteristic boundary conditions* for subcritical and supercritical flows governed by (1.1). Such conditions were first used, to our knowledge, by Nycander *et al.*, [NMF], in numerical experiments with finite difference discretizations of shallow water models. In the paper at hand we will analyze Galerkin-finite element approximations for smooth solutions of the initial-boundary-value problems (ibvp’s) resulting from the application of characteristic boundary conditions to (1.1). The well-posedness of these ibvp’s was analyzed in Petcu & Temam, [PT1] and Huang *et al.*, [HPT].

The characteristic boundary conditions may be derived as follows. We write the system (1.1) as

$$\begin{pmatrix} \eta_t \\ u_t \end{pmatrix} + A \begin{pmatrix} \eta_x \\ u_x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

2010 *Mathematics Subject Classification.* 65M60, 35L60.

Key words and phrases. Shallow water equations, characteristic boundary conditions, Galerkin methods, error estimates.

Work supported by the project PEFYKA of the action KRIPIS of GSRT at IACM/FORTH. The project is funded by Greece and the European Development Fund of EU under the NSRF and the O.P. Competitiveness and Entrepreneurship.

The matrix $A = \begin{pmatrix} u & 1+\eta \\ 1 & u \end{pmatrix}$ has eigenvalues $\lambda_1 = u + \sqrt{1+\eta}$, $\lambda_2 = u - \sqrt{1+\eta}$. Assuming always that $\eta > -1$, we consider two types of flows:

$$\text{Supercritical : } u > \sqrt{1+\eta}, \quad 0 < \lambda_2 < \lambda_1,$$

$$\text{Subcritical : } u < \sqrt{1+\eta}, \quad \lambda_2 < 0 < \lambda_1.$$

It is well known, cf. e.g. [Wh], that along the family of *characteristic curves* with $\dot{x}(t) = \lambda_1 = u + \sqrt{1+\eta}$, the quantity $r_1 := u + 2\sqrt{1+\eta}$ is constant, while along the curves with $\dot{x}(t) = \lambda_2 = u - \sqrt{1+\eta}$, $r_2 := u - 2\sqrt{1+\eta}$ is preserved. If η and u are expressed in terms of the *Riemann invariants* r_1, r_2 one obtains the system of ordinary differential equations (ode's)

$$\begin{aligned} \frac{dr_1}{dt} &= 0 \quad \text{on} \quad x(t) : \quad \frac{dx}{dt} = \lambda_1(r_1, r_2) = \frac{3r_1 + r_2}{4}, \\ \frac{dr_2}{dt} &= 0 \quad \text{on} \quad x(t) : \quad \frac{dx}{dt} = \lambda_2(r_1, r_2) = \frac{r_1 + 3r_2}{4}, \end{aligned}$$

which is equivalent to the original pde system (1.1) and whose discretization yields the classical *method of characteristics* for solving (1.1). If we pose (1.1) in the spatial interval $[0, L]$ it is straightforward to see, cf. [Wh], Section 5.4, that the temporal integration of the ode system requires in the supercritical case that $r_1(0, t)$ and $r_2(0, t)$ be given for $t \geq 0$, as both families of characteristics are incoming at $x = 0$; this is equivalent to prescribing $u(0, t)$ and $\eta(0, t)$ for $t \geq 0$. In the subcritical case $r_1(0, t)$ and $r_2(L, t)$ should be given for $t \geq 0$, since they correspond to the incoming characteristics at $x = 0$ and at $x = L$ respectively.

Following [NMF] we assume that outside the interval $[0, L]$ the flow is uniform and is given by $\eta(x, t) = \eta_0$, $u(x, t) = u_0$, where η_0, u_0 are known constants. Therefore, in the *supercritical case* the characteristic boundary conditions are simply

$$\eta(0, t) = \eta_0, \quad u(0, t) = u_0, \quad (1.3)$$

with $u_0 > \sqrt{1+\eta_0}$, while, in the *subcritical case*, they are of the form

$$\begin{aligned} u(0, t) + 2\sqrt{1+\eta(0, t)} &= u_0 + 2\sqrt{1+\eta_0}, \\ u(L, t) - 2\sqrt{1+\eta(L, t)} &= u_0 - 2\sqrt{1+\eta_0}, \end{aligned} \quad (1.4)$$

where now it is assumed that $u_0^2 < 1 + \eta_0$. In both cases we may view the solution (η, u) of (1.1) for $0 \leq x \leq L$, $t \geq 0$, generated by the initial conditions (1.2), as a perturbation of the uniform flow (η_0, u_0) to which the solution inside the computational domain $[0, L]$ will revert once the waves generated by the initial conditions exit this interval. It is straightforward to check, using the definitions of characteristics and Riemann invariants and considering e.g. initial conditions that differ from η_0, u_0 in a subinterval of $[0, L]$, that the boundary conditions (1.3) and (1.4) are transparent.

As previously mentioned, the characteristic boundary conditions (1.3) and (1.4) were used by Nycander *et al.*, [NMF], in finite difference simulations of the shallow water equations, in one space dimension, actually in more complicated instances of hydraulic and geophysical interest, including single- and two-layer flows in channels of variable width and variable bottom topography, time-dependent forcing in the boundary conditions, examples where transcritical flows develop, *et al.* (In the case of two-layer flows an approximate SW system was used in which the barotropic and baroclinic modes are decoupled; this allows using the analogs of the (local) characteristic boundary conditions in this case too. Also, for reasons of numerical stability, a diffusive term with a small viscosity coefficient was added in the momentum equations.) The finite difference spatial discretization was effected on a staggered grid and the leap-frog scheme was used for time stepping. Similar model equations and characteristic boundary conditions were applied in simulations of two-layer hydraulic exchange flows in [FM].

The characteristic boundary conditions (1.3) and (1.4) were also used by Shiue *et al.*, [SLTT], in the case of the one-dimensional, single-layer shallow water equations in channels of variable bottom topography in the presence of Coriolis terms and with the addition of a cross-velocity variable that depends only on x . The system, written in balance law form, was discretized in space using midpoint quadrature for the source cell integral and a ‘central-upwind’, [KNP], [KP], Godunov-type approximation of the flux term; a second-order, explicit Runge-Kutta method was used for time stepping. Many numerical experiments performed with this scheme are reported in [SLTT]; they simulate interesting cases of subcritical, transcritical and supercritical

flows over variable-bottom topographies. The characteristic boundary conditions and the same numerical scheme were subsequently used in [BPSTT] in the case of two-layer problems in one dimension under the decoupling assumptions of [NMF]. (The local well-posedness of this two-layer problem was studied in [PT2].)

In case the elevation of the free surface η is a small perturbation of the steady state η_0 one may derive *linearized* approximations to the characteristic boundary conditions (1.4) in the subcritical case. These linearized conditions are also considered in [NMF] and in [SLTT], where they are compared to the nonlinear exact conditions and found in general to cause spurious reflections that enter the computational domain. (The linearized boundary conditions are easily seen to be (exactly) transparent for the *linearized* shallow water equations obtained by linearizing (1.1) about the steady state (η_0, u_0) . In [SLTT] it is shown that the ibvp for the linearized system supplemented by the linearized boundary conditions is well posed.) As pointed out in [NMF], [SLTT], and in [ND], the linearized characteristic boundary conditions have been extensively used in the computational fluid dynamics literature; the last reference contains a review of several other absorbing boundary conditions for the shallow water equations at artificial boundaries, including absorbing ('sponge') layer conditions *et al.*

Of particular interest for our purposes is the rigorous analysis of the well-posedness of the ibvp's (locally in time) for the shallow water equations with characteristic boundary conditions carried out in Huang *et al.*, [HPT], and Petcu & Temam, [PT1]. The ibvp in the supercritical case, was studied by Huang *et al.*, [HPT], in fact in the more general setting of shallow water supercritical flows over a variable bottom in the presence of Coriolis terms and a lateral component of the horizontal velocity depending on x , and also with nonhomogeneous boundary conditions satisfying appropriate compatibility conditions. The hypotheses of [HPT] on u_0 , η_0 and the initial data, briefly reviewed in section 2 in the sequel, guarantee the existence and uniqueness, locally in time, of a smooth solution of the ibvp (1.1)-(1.3) with positive $1 + \eta$, satisfying the strong supercriticality property $u^2 - (1 + \eta) \geq c_0^2$ for some positive constant c_0 . The well-posedness of the ibvp in the subcritical case, i.e. of the ibvp (1.1), (1.2), (1.4), was studied by Petcu & Temam, [PT1]. The assumptions of [PT1] (reviewed in section 3 below) imply the existence and uniqueness, locally in time, of a smooth solution of the ibvp with positive $1 + \eta$ and satisfying the strong subcriticality condition $u^2 - (1 + \eta) \leq -c_0^2$, where c_0 is a positive constant.

In this paper we will analyze standard Galerkin-finite element spatial discretizations of the ibvp's (1.1)-(1.3) and (1.1), (1.2), (1.4), under the hypothesis that they have smooth solutions. In both cases the basic approximation will be effected by C^{r-2} functions which are piecewise polynomials of degree $r - 1$, $r \geq 2$, on quasiuniform partitions of $[0, L]$. In section 2 we consider the supercritical case and prove an L^2 -error estimate of $O(h^{r-1})$ accuracy for the Galerkin approximations of η and u . (It is well known that this is the expected best order of convergence in L^2 for standard Galerkin semidiscretizations of first-order hyperbolic problems on general quasiuniform meshes. For *uniform meshes*, better results hold, cf. [D1] for the analysis in the case of a linear model problem. In [AD2] it was proved that the order of convergence in L^2 for piecewise linear continuous elements on a uniform mesh is equal to 2 in the case of an ibvp for (1.1) with the homogeneous boundary conditions $u(0, t) = u(L, t) = 0$. This superaccuracy result is expected to hold for the ibvp's under consideration as well and this is indeed what the numerical experiments of section 4 indicate.) For the proof of the error estimate we assume that a strengthened supercriticality condition holds for the solution of (1.1)-(1.3); cf. (H1)-(H3) in section 2. The proof also requires that $r \geq 3$ so that a certain bootstrap argument, based on the boundedness of the $\|\cdot\|_{1,\infty}$ norm of an error term, goes through as in [D2], [AD2].

In section 3 we turn to the subcritical case. We write the ibvp (1.1), (1.2), (1.4) in its classical diagonal form in which the new unknowns are analogs of the two Riemann invariants in the context of the ibvp at hand and satisfy homogeneous Dirichlet boundary conditions one at $x = 0$ and the other at $x = L$. The diagonal system is discretized in space on a quasiuniform mesh by the same type of standard Galerkin method as before, and a L^2 error estimate of $O(h^{r-1})$ is proved for both components of the solution. A change of variables of this semidiscrete approximation yields approximations of the original unknowns η and u of $O(h^{r-1})$ accuracy. The proof requires that a strengthened form of the subcriticality property holds for the solution of the ibvp (1.1), (1.2), (1.4) (cf. (Y1), (Y2) in section 3), and the technical assumption that $r \geq 3$.

Section 4 is a report of various numerical experiments that we performed with the Galerkin-finite element methods of sections 2 and 3 and some of their variants. We use spatial discretizations with piecewise linear

continuous functions on uniform meshes and discretize them in the temporal variable by the ‘classical’, explicit, four-stage, fourth-order Runge-Kutta scheme. The resulting fully discrete methods are stable under a Courant number restriction. (Stability and convergence of high order explicit Runge-Kutta methods was established for closely related pde systems in [AD1] and [AD2].) Our main purpose in the numerical experiments is to check the stability and the numerical order of convergence of the fully discrete Galerkin methods and study by computational means their absorption properties. Although the full discretizations of the characteristic boundary conditions are not exactly transparent of course, the numerical experiments show that they are practically transparent, in contrast to the analogous Galerkin schemes with the linearized boundary conditions that we also implement in the subcritical case; the latter are absorbing but in general allow spurious reflections to form and enter the computational domain. In the subcritical case we also implement the analogous fully discrete Galerkin method for the original, nondiagonal form (1.1), (1.2), (1.4) of the system and check that it gives results close but somewhat inferior to those of the analogous discretization of the diagonal form of the system.

In the error estimates in the sequel, we let the spatial interval be $[0, 1]$ for simplicity. We let $C^k = C^k[0, 1]$, $k = 0, 1, 2, \dots$, be the space of k times continuously differentiable functions on $[0, 1]$. The norm and inner product on $L^2 = L^2(0, 1)$ are denoted by $\|\cdot\|$, (\cdot, \cdot) , respectively. For integer $k \geq 0$, H^k , $\|\cdot\|_k$ will denote the usual, L^2 -based Sobolev spaces of classes of functions and the associated norms. The norms on $L^\infty = L^\infty(0, 1)$ and on the L^∞ -based Sobolev spaces $W_\infty^k = W_\infty^k(0, 1)$ will be denoted by $\|\cdot\|_\infty$, $\|\cdot\|_{k,\infty}$, respectively. Finally, we let \mathbb{P}_r be the polynomials of degree $\leq r$.

2. SEMIDISCRETIZATION OF THE SUPERCRITICAL SHALLOW WATER EQUATIONS

In this section we consider the shallow water equations with characteristic boundary conditions in the *supercritical* case. Specifically, for $(x, t) \in [0, 1] \times [0, T]$ we seek $\eta = \eta(x, t)$ and $u = u(x, t)$ satisfying the ibvp

$$\begin{aligned} \eta_t + u_x + (\eta u)_x &= 0, \\ u_t + \eta_x + uu_x &= 0, \\ \eta(x, 0) &= \eta^0(x), \quad u(x, 0) = u^0(x), \quad 0 \leq x \leq 1, \\ \eta(0, t) &= \eta_0, \quad u(0, t) = u_0, \quad 0 \leq t \leq T, \end{aligned} \tag{SW1}$$

where η^0, u^0 are given functions on $[0, 1]$ and η_0, u_0 constants such that $1 + \eta_0 > 0$, $u_0 > 0$, $u_0 > \sqrt{1 + \eta_0}$.

As mentioned in the Introduction, the ibvp (SW1) was studied by Huang *et al.*, [HPT], in fact in the more general case of a shallow water supercritical flow with nonhomogeneous boundary conditions over a variable bottom for a nonzero Coriolis parameter and also in the presence of a lateral component of the horizontal velocity depending on x only. In the simpler case of (SW1), the proof of the main result of [HPT] amounts to the selection of a suitable constant solution (η_0, u_0) of (SW1) and of sufficiently smooth initial conditions close to the constant solution and satisfying appropriate compatibility relations at $x = 0$. Under these hypotheses the conclusion of [HPT] is that given positive constants $c_0, \alpha_0, \underline{\zeta}_0$, and $\bar{\zeta}_0$, there exists a $T > 0$ such that a sufficiently smooth solution of (SW1) exists satisfying for $(x, t) \in [0, 1] \times [0, T]$ the strong supercriticality properties

$$u^2 - (1 + \eta) \geq c_0^2, \tag{P1}$$

$$u \geq \alpha_0, \tag{P2}$$

$$\underline{\zeta}_0 \leq (1 + \eta) \leq \bar{\zeta}_0. \tag{P3}$$

For the purposes of the error estimation to follow we will assume that (SW1) has a sufficiently smooth solution (η, u) that satisfies a strengthened supercriticality condition of the following form: There exist positive constants α and β , such that for $(x, t) \in [0, 1] \times [0, T]$ it holds that

$$1 + \eta \geq \beta, \tag{H1}$$

$$u \geq 2\alpha, \tag{H2}$$

$$1 + \eta \leq (u - \alpha)(u - \frac{2\alpha}{3}). \tag{H3}$$

Obviously (H1) and (H3) imply that $u > \sqrt{1+\eta}$. It is not hard to see that (H3) follows from (P1)-(P3) if e.g. α_0 is taken sufficiently small and c_0 sufficiently large. We also remark here that in the error estimates to follow (H3) will be needed only at $x = 1$ for $t \in [0, T]$.

We will approximate the solution of (SW1) in a slightly tranformed form. We let $\tilde{\eta} = \eta - \eta_0$, $\tilde{u} = u - u_0$ and rewrite (SW1) as an ibvp for $\tilde{\eta}$ and \tilde{u} with homogeneous boundary conditions. Dropping the tildes we obtain the system

$$\begin{aligned} \eta_t + u_0\eta_x + (1 + \eta_0)u_x + (\eta u)_x &= 0, \\ u_t + \eta_x + u_0u_x + uu_x &= 0, \\ \eta(x, 0) &= \eta^0(x) - \eta_0, \quad u(x, 0) = u^0(x) - u_0, \quad 0 \leq x \leq 1, \\ \eta(0, t) &= 0, \quad u(0, t) = 0, \quad 0 \leq t \leq T. \end{aligned} \tag{SW1a}$$

In terms of the new variables η and u , our hypotheses (H1)-(H3) become

$$1 + \eta + \eta_0 \geq \beta, \tag{H1a}$$

$$u + u_0 \geq 2\alpha, \tag{H2a}$$

$$1 + \eta + \eta_0 \leq (u + u_0 - \alpha)(u + u_0 - \frac{2\alpha}{3}). \tag{H3a}$$

In the sequel, for integer $k \geq 0$, let $\mathring{C}^k = \{v \in C^k[0, 1] : v(0) = 0\}$, and $\mathring{H}^{k+1} = \{v \in H^{k+1}(0, 1) : v(0) = 0\}$. For a positive integer N let $0 = x_1 < x_2 < \dots < x_{N+1} = 1$ be a quasiuniform partition of $[0, 1]$ with $h := \max_i(x_{i+1} - x_i)$, and for integer $r \geq 2$ define $\mathring{S}_h = \{\phi \in \mathring{C}^{r-2} : \phi|_{[x_j, x_{j+1}]} \in \mathbb{P}_{r-1}, 1 \leq j \leq N\}$. It is well known that if $v \in \mathring{H}^r$, there exists $\chi \in \mathring{S}_h$ such that

$$\|v - \chi\| + h\|v' - \chi'\| \leq Ch^r\|v\|_r, \tag{2.1}$$

and, cf. [Sch], if $r \geq 3$,

$$\|v - \chi\|_2 \leq Ch^{r-2}\|v\|_r. \tag{2.2}$$

(Here and in the sequel C will denote a generic constant independent of h .) In addition, if P is the L^2 -projection operator onto \mathring{S}_h , then it follows that, cf. [DDW],

$$\|Pv\|_\infty \leq C\|v\|_\infty, \quad \text{if } v \in L^\infty, \tag{2.3}$$

$$\|Pv - v\|_\infty \leq Ch^r\|v\|_{r,\infty}, \quad \text{if } v \in W^{r,\infty} \cap \mathring{H}^1. \tag{2.4}$$

As a consequence of the quasiuniformity of the mesh the inverse inequalities

$$\|\chi\|_1 \leq Ch^{-1}\|\chi\|, \tag{2.5}$$

$$\|\chi\|_{j,\infty} \leq Ch^{-(j+1/2)}\|\chi\|, \quad j = 0, 1, \tag{2.6}$$

hold for $\chi \in \mathring{S}_h$. (In (2.6) $\|\cdot\|_{0,\infty} = \|\cdot\|_\infty$.)

The standard Galerkin semidiscretization of (SW1a) is defined as follows: We seek $\eta_h, u_h : [0, T] \rightarrow \mathring{S}_h$ such that for $0 \leq t \leq T$

$$(\eta_{ht}, \phi) + (u_0\eta_{hx}, \phi) + ((1 + \eta_0)u_{hx}, \phi) + ((\eta_h u_h)_x, \phi) = 0, \quad \forall \phi \in \mathring{S}_h, \tag{2.7}$$

$$(u_{ht}, \phi) + (\eta_{hx}, \phi) + (u_0u_{hx}, \phi) + (u_h u_{hx}, \phi) = 0, \quad \forall \phi \in \mathring{S}_h, \tag{2.8}$$

with

$$\eta_h(0) = P(\eta^0(\cdot) - \eta_0), \quad u_h(0) = P(u^0(\cdot) - u_0). \tag{2.9}$$

The main result of this section is:

Proposition 2.1. *Let (η, u) be the solution of (SW1a), and assume that the hypotheses (H1a), (H2a), (H3a) hold, that $r \geq 3$, and h is sufficiently small. Then the semidiscrete ivp (2.7)-(2.9) has a unique solution (η_h, u_h) for $0 \leq t \leq T$ satisfying*

$$\max_{0 \leq t \leq T} (\|\eta(t) - \eta_h(t)\| + \|u(t) - u_h(t)\|) \leq Ch^{r-1}. \tag{2.10}$$

Proof. Let $\rho = \eta - P\eta$, $\theta = P\eta - \eta_h$, $\sigma = u - Pu$, $\xi = Pu - u_h$. After choosing a basis for \mathring{S}_h , it is straightforward to see that the semidiscrete problem (2.7)-(2.9) represents an ivp for an ode system which has a unique solution locally in time. While this solution exists, it follows from (2.7), (2.8) and the pde's in (SW1a), that

$$(\theta_t, \phi) + (u_0(\rho_x + \theta_x), \phi) + ((1 + \eta_0)(\sigma_x + \xi_x), \phi) + ((\eta u - \eta_h u_h)_x, \phi) = 0, \quad \forall \phi \in \mathring{S}_h, \quad (2.11)$$

$$(\xi_t, \phi) + (\rho_x + \theta_x, \phi) + (u_0(\sigma_x + \xi_x), \phi) + (uu_x - u_h u_{hx}, \phi) = 0, \quad \forall \phi \in \mathring{S}_h. \quad (2.12)$$

Since $\eta u - \eta_h u_h = \eta(\sigma + \xi) + u(\rho + \theta) - (\rho + \theta)(\sigma + \xi)$, $uu_x - u_h u_{hx} = (u\sigma)_x + (u\xi)_x - (\sigma\xi)_x - \sigma\sigma_x - \xi\xi_x$, it follows that

$$(\eta u - \eta_h u_h)_x = (\eta\xi)_x + (u\theta)_x - (\theta\xi)_x + \tilde{R}_1, \quad (2.13)$$

$$uu_x - u_h u_{hx} = (u\xi)_x - \xi\xi_x + \tilde{R}_2, \quad (2.14)$$

where

$$\tilde{R}_1 = (\eta\sigma)_x + (u\rho)_x - (\rho\sigma)_x - (\rho\xi)_x - (\theta\sigma)_x, \quad (2.15)$$

$$\tilde{R}_2 = (u\sigma)_x - (\sigma\xi)_x - \sigma\sigma_x. \quad (2.16)$$

Therefore, the equations (2.11), (2.12) may be written as

$$(\theta_t, \phi) + (u_0\theta_x, \phi) + (\gamma_x, \phi) + ((u\theta)_x, \phi) - ((\theta\xi)_x, \phi) = -(R_1, \phi), \quad \forall \phi \in \mathring{S}_h, \quad (2.17)$$

$$(\xi_t, \phi) + (\theta_x, \phi) + (u_0\xi_x, \phi) + ((u\xi)_x, \phi) - (\xi\xi_x, \phi) = -(R_2, \phi), \quad \forall \phi \in \mathring{S}_h. \quad (2.18)$$

where $\gamma = (1 + \eta_0 + \eta)\xi$ and

$$R_1 = u_0\rho_x + (1 + \eta_0)\sigma_x + \tilde{R}_1, \quad (2.19)$$

$$R_2 = \rho_x + u_0\sigma_x + \tilde{R}_2. \quad (2.20)$$

Putting $\phi = \theta$ in (2.17), using integration by parts, and suppressing the dependence on t we have

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \|\theta\|^2 - (\gamma, \theta_x) + \frac{1}{2} (u_0 + u(1))\theta^2(1) + (1 + \eta_0 + \eta(1))\xi(1)\theta(1) \\ - \frac{1}{2} \xi(1)\theta^2(1) = -\frac{1}{2} (u_x\theta, \theta) + \frac{1}{2} (\xi_x\theta, \theta) - (R_1, \theta). \end{aligned} \quad (2.21)$$

Take now $\phi = P\gamma = P[(1 + \eta_0 + \eta)\xi]$ in (2.18) and get

$$(\xi_t, \gamma) + (\theta_x, \gamma) + (u_0\xi_x, \gamma) + ((u\xi)_x, \gamma) - (\xi\xi_x, \gamma) = -(R_3, P\gamma - \gamma) - (R_2, P\gamma), \quad (2.22)$$

where

$$R_3 = \theta_x + u_0\xi_x + (u\xi)_x - \xi\xi_x. \quad (2.23)$$

Integration by parts in various terms in (2.22) gives

$$\begin{aligned} (u_0\xi_x, \gamma) &= (u_0\xi_x, (1 + \eta_0 + \eta)\xi) = \frac{1}{2} u_0(1 + \eta_0 + \eta(1))\xi^2(1) - \frac{1}{2} (u_0\eta_x\xi, \xi), \\ ((u\xi)_x, \gamma) &= ((u\xi)_x, (1 + \eta_0 + \eta)\xi) = (u_x\xi, (1 + \eta_0 + \eta)\xi) + (u\xi_x, (1 + \eta_0 + \eta)\xi) \\ &= \frac{1}{2} u(1)(1 + \eta_0 + \eta(1))\xi^2(1) + \frac{1}{2} (u_x(1 + \eta_0 + \eta), \xi^2) - \frac{1}{2} (u\eta_x\xi, \xi), \\ (\xi\xi_x, \gamma) &= (\xi\xi_x, (1 + \eta_0 + \eta)\xi) = \frac{1}{3} (1 + \eta_0 + \eta(1))\xi^3(1) - \frac{1}{3} (\eta_x\xi^2, \xi). \end{aligned}$$

Hence (2.22) becomes

$$\begin{aligned} (\xi_t, \gamma) + (\theta_x, \gamma) + \frac{1}{2} (u_0 + u(1))(1 + \eta_0 + \eta(1))\xi^2(1) - \frac{1}{3} (1 + \eta_0 + \eta(1))\xi^3(1) \\ = (R_4, \xi) - (R_3, P\gamma - \gamma) - (R_2, P\gamma), \end{aligned} \quad (2.24)$$

where

$$R_4 = \frac{1}{2} u_0\eta_x\xi - \frac{1}{2} u_x(1 + \eta_0 + \eta)\xi + \frac{1}{2} u\eta_x\xi - \frac{1}{3} \eta_x\xi^2. \quad (2.25)$$

Adding now (2.21) and (2.24) we obtain

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} [\|\theta\|^2 + ((1 + \eta_0 + \eta)\xi, \xi)] + \omega = \frac{1}{2} (\eta_t\xi, \xi) - \frac{1}{2} (u_x\theta, \theta) \\ + \frac{1}{2} (\xi_x\theta, \theta) - (R_1, \theta) + (R_4, \xi) - (R_3, P\gamma - \gamma) - (R_2, P\gamma), \end{aligned} \quad (2.26)$$

where

$$\begin{aligned} \omega = & \frac{1}{2}(u_0 + u(1))\theta^2(1) + \frac{1}{2}(u_0 + u(1))(1 + \eta_0 + \eta(1))\xi^2(1) \\ & + (1 + \eta_0 + \eta(1))\xi(1)\theta(1) - \frac{1}{2}\xi(1)\theta^2(1) - \frac{1}{3}(1 + \eta_0 + \eta(1))\xi^3(1). \end{aligned} \quad (2.27)$$

In view of (2.9), by continuity we conclude that there exists a maximal temporal instance $t_h > 0$ such that (η_h, u_h) exist and $\|\xi_x\|_\infty \leq \alpha$ for $t \leq t_h$. Suppose that $t_h < T$. Then, since $\|\xi\|_\infty \leq \|\xi_x\|_\infty$, it follows from (2.27) that for $t \in [0, t_h]$

$$\begin{aligned} \omega \geq & \frac{1}{2}(u_0 + u(1) - \alpha)\theta^2(1) + \frac{1}{2}(1 + \eta_0 + \eta(1))(u_0 + u(1) - \frac{2\alpha}{3})\xi^2(1) \\ & + (1 + \eta_0 + \eta(1))\xi(1)\theta(1) = \frac{1}{2}(\theta(1), \xi(1))^T \begin{pmatrix} \mu & \lambda \\ \lambda & \lambda\nu \end{pmatrix} \begin{pmatrix} \theta(1) \\ \xi(1) \end{pmatrix}, \end{aligned} \quad (2.28)$$

where $\mu = u_0 + u(1) - \alpha$, $\lambda = 1 + \eta_0 + \eta(1)$, $\nu = u_0 + u(1) - \frac{2\alpha}{3}$. The hypotheses (H1a) and (H2a) give that $0 < \mu < \nu$, $\lambda > 0$. It is easy to see then that the matrix in (2.28) will be positive semidefinite precisely when (H3a) holds. We conclude from (2.28) that $\omega \geq 0$.

We now estimate the various terms in the right-hand side of (2.26) for $0 \leq t \leq t_h$. We obviously have

$$|(\eta_t \xi, \xi)| + |(u_x \theta, \theta)| \leq C(\|\xi\|^2 + \|\theta\|^2), \quad (2.29)$$

and

$$|(\xi_x \theta, \theta)| \leq \alpha \|\theta\|^2. \quad (2.30)$$

In addition, from (2.19), (2.14), and the inverse and approximation properties of \mathring{S}_h and (2.3), (2.4) we have

$$\begin{aligned} |(R_1, \theta)| & \leq Ch^{r-1} \|\theta\| + \|\rho\|_\infty \|\xi_x\| \|\theta\| + \|\rho_x\|_\infty \|\xi\| \|\theta\| \\ & \quad + \|\sigma_x\|_\infty \|\theta\|^2 + \|\sigma\|_\infty \|\theta_x\| \|\theta\| \\ & \leq Ch^{r-1} \|\theta\| + C(\|\theta\|^2 + \|\xi\|^2). \end{aligned} \quad (2.31)$$

Also, from (2.25)

$$|(R_4, \xi)| \leq C\|\xi\|^2 + C\|\xi\|_\infty \|\xi\|^2 \leq C(1 + \alpha)\|\xi\|^2. \quad (2.32)$$

By (2.20), (2.16), (2.3), (2.4) and the inverse and approximation properties of \mathring{S}_h we have

$$\begin{aligned} |(R_2, P\gamma)| & \leq Ch^{r-1} \|\xi\| + C\|\sigma\|_\infty \|\xi_x\| \|\xi\| + C\|\sigma_x\|_\infty \|\xi\|^2 \\ & \leq Ch^{r-1} \|\xi\| + C\|\xi\|^2. \end{aligned} \quad (2.33)$$

Finally, using a well-known *superapproximation* property of \mathring{S}_h , cf. [DDW], [D2], in order to estimate the term $P\gamma - \gamma$ by

$$\|P\gamma - \gamma\| = \|P[(1 + \eta + \eta_0)\xi] - (1 + \eta + \eta_0)\xi\| \leq Ch\|\xi\|, \quad (2.34)$$

we obtain by (2.23) and the inverse properties of \mathring{S}_h that

$$\begin{aligned} |(R_3, P\gamma - \gamma)| & \leq |(\theta_x, P\gamma - \gamma)| + |(u_0 \xi_x, P\gamma - \gamma)| + |((u\xi)_x, P\gamma - \gamma)| + |(\xi \xi_x, P\gamma - \gamma)| \\ & \leq Ch\|\theta_x\| \|\xi\| + Ch\|\xi_x\| \|\xi\| + Ch\|\xi\|^2 + Ch\|\xi\|_\infty \|\xi_x\| \|\xi\| \\ & \leq C\|\theta\| \|\xi\| + C\|\xi\|^2 + C\alpha\|\xi\|^2 \\ & \leq C\|\theta\|^2 + C(1 + \alpha)\|\xi\|^2. \end{aligned} \quad (2.35)$$

Therefore, (2.26), the fact that $\omega \geq 0$, and the inequalities (2.30)-(2.33), (2.35) give for $0 \leq t \leq t_h$

$$\frac{d}{dt} [\|\theta\|^2 + ((1 + \eta_0 + \eta)\xi, \xi)] \leq Ch^{r-1} (\|\theta\| + \|\xi\|) + C(\|\theta\|^2 + \|\xi\|^2),$$

where C is a constant independent of h and t_h . By (H1a) the norm $((1 + \eta_0 + \eta)\cdot, \cdot)^{1/2}$ is equivalent to that of L^2 uniformly for $t \in [0, T]$. Hence, Gronwall's inequality and the fact that $\theta(0) = \xi(0) = 0$ yield for a constant $C = C(T)$

$$\|\theta\| + \|\xi\| \leq Ch^{r-1} \quad \text{for } 0 \leq t \leq t_h. \quad (2.36)$$

We conclude from (2.6) that $\|\xi_x\|_\infty \leq Ch^{r-5/2}$ for $0 \leq t \leq t_h$, and, since $r \geq 3$, if h is taken sufficiently small, t_h is not maximal. Hence we may take $t_h = T$ and (2.10) follows from (2.36). \square

3. SEMIDISCRETIZATION OF THE SUBCRITICAL SHALLOW WATER EQUATIONS

In this section we consider the shallow water equations with characteristic boundary conditions in the *subcritical* case. Specifically, for $(x, t) \in [0, 1] \times [0, T]$ we seek $\eta = \eta(x, t)$ and $u = u(x, t)$ satisfying the ibvp

$$\begin{aligned} \eta_t + u_x + (\eta u)_x &= 0, \\ u_t + \eta_x + uu_x &= 0, \\ \eta(x, 0) &= \eta^0(x), \quad u(x, 0) = u^0(x), \quad 0 \leq x \leq 1, \\ u(0, t) + 2\sqrt{1 + \eta(0, t)} &= u_0 + 2\sqrt{1 + \eta_0}, \quad 0 \leq t \leq T, \\ u(1, t) - 2\sqrt{1 + \eta(1, t)} &= u_0 - 2\sqrt{1 + \eta_0}, \quad 0 \leq t \leq T, \end{aligned} \tag{SW2}$$

where η^0, u^0 are given functions on $[0, 1]$ and η_0, u_0 constants such that $1 + \eta_0 > 0$ and $u_0^2 < 1 + \eta_0$.

As mentioned in the Introduction, the ibvp (SW2) was studied by Petcu & Temam, [PT1]. They used the hypotheses that there exists a constant $c_0 > 0$ such that $u_0^2 - (1 + \eta_0) \leq -c_0^2$ and that the initial conditions $\eta^0(x)$ and $u^0(x)$ are sufficiently smooth and satisfy the condition $(u^0(x))^2 - (1 + \eta^0(x)) \leq -c_0^2$ (with $1 + \eta^0(x)$ positive) and suitable compatibility relations at $x = 0$ and $x = 1$. Under these assumptions one may infer from the theory of [PT1] that there exists a $T > 0$ such that a sufficiently smooth solution (η, u) of (SW2) exists for $(x, t) \in [0, 1] \times [0, T]$ with the properties that $1 + \eta$ is positive and the strong subcriticality condition

$$u^2 - (1 + \eta) \leq -c_0^2, \tag{II}$$

holds for $(x, t) \in [0, 1] \times [0, T]$. For the purposes of the error estimation to follow we will assume that (SW2) has a sufficiently smooth solution (η, u) such that $1 + \eta > 0$ and satisfies a stronger subcriticality condition. Specifically we assume that for some constant $c_0 > 0$ it holds that

$$u_0 + \sqrt{1 + \eta_0} \geq c_0, \quad u_0 - \sqrt{1 + \eta_0} \leq -c_0, \tag{Y1}$$

and for $(x, t) \in [0, 1] \times [0, T]$ that

$$u + \sqrt{1 + \eta} \geq c_0, \quad u - \sqrt{1 + \eta} \leq -c_0. \tag{Y2}$$

Obviously (Y1) and (Y2) imply the subcriticality conditions $u_0^2 - (1 + \eta_0) \leq -c_0^2$ and (II) of [PT1], and approximate the latter better as c_0 decreases.

In this section we will approximate the solution of (SW2) with a Galerkin-finite element method after transforming (1.1) in its classical diagonal form. As in the Introduction, we write the system as

$$\begin{pmatrix} \eta_t \\ u_t \end{pmatrix} + A \begin{pmatrix} \eta_x \\ u_x \end{pmatrix} = 0, \tag{3.1}$$

where $A = \begin{pmatrix} u & 1 + \eta \\ 1 & u \end{pmatrix}$. The matrix A has the eigenvalues $\lambda_1 = u + \sqrt{1 + \eta}$, $\lambda_2 = u - \sqrt{1 + \eta}$, (note that by (Y2) $\lambda_1 \geq c_0$ and $\lambda_2 \leq -c_0$ in $[0, 1] \times [0, T]$), with associated eigenvectors $X_1 = (\sqrt{1 + \eta}, 1)^T$, $X_2 = (-\sqrt{1 + \eta}, 1)^T$. If S is the matrix with columns X_1, X_2 it follows from (3.1) that

$$S^{-1} \begin{pmatrix} \eta_t \\ u_t \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} S^{-1} \begin{pmatrix} \eta_x \\ u_x \end{pmatrix} = 0. \tag{3.2}$$

If we try to define now functions v, w on $[0, 1] \times [0, T]$ by the equations $S^{-1} \begin{pmatrix} \eta_t \\ u_t \end{pmatrix} = \begin{pmatrix} v_t \\ w_t \end{pmatrix}$, $S^{-1} \begin{pmatrix} \eta_x \\ u_x \end{pmatrix} = \begin{pmatrix} v_x \\ w_x \end{pmatrix}$, we see that these equations are consistent and their solutions are given by $v = \frac{1}{2}(u + 2\sqrt{1 + \eta}) + c_v$, $w = \frac{1}{2}(u - 2\sqrt{1 + \eta}) + c_w$, for arbitrary constants c_v, c_w . Choosing the constants c_v, c_w so that $v(0, t) = 0$, $w(1, t) = 0$, and using the boundary conditions in (SW2) we get

$$v = \frac{1}{2}[u - u_0 + 2(\sqrt{1 + \eta} - \delta_0)], \quad w = \frac{1}{2}[u - u_0 - 2(\sqrt{1 + \eta} - \delta_0)], \tag{3.3}$$

where $\delta_0 = \sqrt{1 + \eta_0}$. The original variables η, u are given in terms of v and w by the formulas

$$\eta = [\tfrac{1}{2}(v - w) + \delta_0]^2 - 1, \quad u = v + w + u_0. \tag{3.4}$$

Since

$$\lambda_1 = u + \sqrt{1 + \eta} = u_0 + \delta_0 + \frac{3v + w}{2}, \quad \lambda_2 = u - \sqrt{1 + \eta} = u_0 - \delta_0 + \frac{v + 3w}{2}, \quad (3.5)$$

we see that the ibvp (SW2) becomes

$$\begin{aligned} \begin{pmatrix} v_t \\ w_t \end{pmatrix} + \begin{pmatrix} u_0 + \delta_0 + \frac{3v+w}{2} & 0 \\ 0 & u_0 - \delta_0 + \frac{v+3w}{2} \end{pmatrix} \begin{pmatrix} v_x \\ w_x \end{pmatrix} &= 0, \quad 0 \leq x \leq 1, \quad 0 \leq t \leq T, \\ v(x, 0) &= v^0(x), \quad w(x, 0) = w^0(x), \quad 0 \leq x \leq 1, \\ v(0, t) &= 0, \quad w(1, t) = 0, \quad 0 \leq t \leq T, \end{aligned} \quad (\text{SW2a})$$

where $v^0(x) = \frac{1}{2}[u^0(x) - u_0 + 2(\sqrt{1 + \eta^0(x)} - \delta_0)]$, $w^0(x) = \frac{1}{2}[u^0(x) - u_0 - 2(\sqrt{1 + \eta^0(x)} - \delta_0)]$. Under our hypotheses (SW2a) has a unique solution (v, w) on $[0, 1] \times [0, T]$ which will be assumed to be smooth enough for the purposes of the error estimation that follows. Of course, v and w represent analogs of the Riemann invariants of the shallow water system in the context of the ibvp at hand; the system of pde's in (SW2a) and (3.4), (3.5) imply that the solution (η, u) of (SW2) may be expressed in terms of two waves v and w that propagate to the right and left, respectively, with speeds $u + \sqrt{1 + \eta}$ and $u - \sqrt{1 + \eta}$.

Given a quasiniform partition of $[0, 1]$ as in section 2, in addition to the spaces defined there, let for integer $k \geq 0$ $\mathring{C}^k = \{f \in C^k[0, 1] : f(1) = 0\}$, $\mathring{H}^{k+1} = \{f \in H^{k+1}(0, 1), f(1) = 0\}$, and, for integer $r \geq 2$, $\mathring{S}_h = \{\phi \in \mathring{C}^{r-2} : \phi|_{[x_j, x_{j+1}]} \in \mathbb{P}_{r-1}, 1 \leq j \leq N\}$. Note that the analogs of the approximation and inverse properties (2.1), (2.2), (2.5), (2.6) hold for \mathring{S}_h as well, and that (2.3), (2.4) are also valid for the L^2 projection \mathcal{P} onto \mathring{S}_h , *mutatis mutandis*.

The (standard) Galerkin semidiscretization of (SW2a) is then defined as follows: Seek $v_h : [0, T] \rightarrow \mathring{S}_h$, $w_h : [0, T] \rightarrow \mathring{S}_h$, such that for $t \in [0, T]$

$$(v_{ht}, \phi) + ((u_0 + \delta_0)v_{hx}, \phi) + \frac{3}{2}(v_h v_{hx}, \phi) + \frac{1}{2}(w_h v_{hx}, \phi) = 0, \quad \forall \phi \in \mathring{S}_h, \quad (3.6)$$

$$(w_{ht}, \chi) + ((u_0 - \delta_0)w_{hx}, \chi) + \frac{3}{2}(w_h w_{hx}, \chi) + \frac{1}{2}(v_h w_{hx}, \chi) = 0, \quad \forall \chi \in \mathring{S}_h, \quad (3.7)$$

with

$$v_h(0) = P(v^0), \quad w_h(0) = \mathcal{P}(w^0). \quad (3.8)$$

The main result of this section is

Proposition 3.1. *Let (v, w) be the solution of (SW2a) and assume that the hypotheses (Y1) and (Y2) hold, that $r \geq 3$, and that h is sufficiently small. Then the semidiscrete ivp (3.6)-(3.8) has a unique solution (v_h, w_h) for $0 \leq t \leq T$ that satisfies*

$$\max_{0 \leq t \leq T} (\|v - v_h\| + \|w - w_h\|) \leq Ch^{r-1}. \quad (3.9)$$

If (η, u) is the solution of (SW2) and we define

$$\eta_h = [\frac{1}{2}(v_h - w_h) + \delta_0]^2 - 1, \quad u_h = v_h + w_h + u_0, \quad (3.10)$$

then

$$\max_{0 \leq t \leq T} (\|\eta - \eta_h\| + \|u - u_h\|) \leq Ch^{r-1}.$$

Proof. Let $\rho = v - Pv$, $\theta = Pv - v_h$, $\sigma = w - \mathcal{P}w$, $\xi = \mathcal{P}w - w_h$. After choosing bases for \mathring{S}_h and \mathring{S}_h we see that the ode ivp (3.6)-(3.8) has a unique solution locally in time. From (SW2a) and (3.6), (3.7) we obtain, as long as the solution exists,

$$(\theta_t, \phi) + ((u_0 + \delta_0)(\theta_x + \rho_x), \phi) + \frac{3}{2}(v v_x - v_h v_{hx}, \phi) + \frac{1}{2}((w v_x - w_h v_{hx}, \phi) = 0, \quad \forall \phi \in \mathring{S}_h, \quad (3.11)$$

$$(\xi_t, \chi) + ((u_0 - \delta_0)(\sigma_x + \xi_x), \chi) + \frac{3}{2}(w w_x - w_h w_{hx}, \chi) + \frac{1}{2}(v w_x - v_h w_{hx}, \chi) = 0, \quad \forall \chi \in \mathring{S}_h. \quad (3.12)$$

Now, since

$$\begin{aligned}
vv_x - v_h v_{hx} &= (v\rho)_x + (v\theta)_x - (\rho\theta)_x - \rho\rho_x - \theta\theta_x, \\
wv_x - w_h v_{hx} &= w(\rho_x + \theta_x) + v_x(\sigma + \xi) - (\rho_x + \theta_x)(\sigma + \xi), \\
ww_x - w_h w_{hx} &= (w\sigma)_x + (w\xi)_x - (\sigma\xi)_x - \sigma\sigma_x - \xi\xi_x, \\
vv_x - v_h w_{hx} &= v(\sigma_x + \xi_x) + w_x(\rho + \theta) - (\sigma_x + \xi_x)(\rho + \theta),
\end{aligned}$$

it follows that

$$vv_x - v_h v_{hx} = (v\theta)_x - \theta\theta_x + R_{11}, \quad wv_x - w_h v_{hx} = -\theta_x\xi + R_{12}, \quad (3.13)$$

$$ww_x - w_h w_{hx} = (w\xi)_x - \xi\xi_x + R_{21}, \quad vv_x - v_h w_{hx} = -\xi_x\theta + R_{22}, \quad (3.14)$$

where

$$R_{11} = (v\rho)_x - (\rho\theta)_x - \rho\rho_x, \quad R_{12} = w\rho_x + w\theta_x + v_x\sigma + v_x\xi - \rho_x\sigma - \rho_x\xi - \theta_x\sigma, \quad (3.15)$$

$$R_{21} = (w\sigma)_x - (\sigma\xi)_x - \sigma\sigma_x, \quad R_{22} = v\sigma_x + v\xi_x + w_x\rho + w_x\theta - \sigma_x\rho - \sigma_x\theta - \xi_x\rho. \quad (3.16)$$

Putting now $\phi = \theta$ in (3.11) and $\chi = \xi$ in (3.12) we obtain

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\theta\|^2 + ((u_0 + \delta_0)\theta_x, \theta) + \frac{3}{2}((v\theta)_x, \theta) - \frac{3}{2}(\theta\theta_x, \theta) \\
= -((u_0 + \delta_0)\rho_x, \theta) - \frac{3}{2}(R_{11}, \theta) + \frac{1}{2}(\theta_x\xi, \theta) - \frac{1}{2}(R_{12}, \theta),
\end{aligned} \quad (3.17)$$

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\xi\|^2 + ((u_0 - \delta_0)\xi_x, \xi) + \frac{3}{2}((w\xi)_x, \xi) - \frac{3}{2}(\xi\xi_x, \xi) \\
= -((u_0 - \delta_0)\sigma_x, \xi) - \frac{3}{2}(R_{21}, \xi) + \frac{1}{2}(\xi_x\theta, \xi) - \frac{1}{2}(R_{22}, \theta),
\end{aligned} \quad (3.18)$$

Integration by parts yields (we suppress the t -dependence)

$$\begin{aligned}
((u_0 + \delta_0)\theta_x, \theta) &= \frac{u_0 + \delta_0}{2} \theta^2(1), \quad ((v\theta)_x, \theta) = \frac{1}{2}(v_x\theta, \theta) + \frac{1}{2}v(1)\theta^2(1), \\
(\theta\theta_x, \theta) &= \frac{1}{3}\theta^3(1), \quad ((u_0 - \delta_0)\xi_x, \xi) = -\frac{u_0 - \delta_0}{2}\xi^2(0), \\
((w\xi)_x, \xi) &= \frac{1}{2}(w_x\xi, \xi) - \frac{1}{2}w(0)\xi^2(0), \quad (\xi\xi_x, \xi) = -\frac{1}{3}\xi^3(0).
\end{aligned}$$

Hence, (3.17) becomes

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\theta\|^2 + \frac{1}{2}(u_0 + \delta_0 + \frac{3}{2}v(1) - \theta(1))\theta^2(1) \\
= -((u_0 + \delta_0)\rho_x, \theta) - \frac{3}{4}(v_x\theta, \theta) + \frac{1}{2}(\theta_x\xi, \theta) - \frac{3}{2}(R_{11}, \theta) - \frac{1}{2}(R_{12}, \theta).
\end{aligned}$$

By (Y2) and (3.5) we see that $u_0 + \delta_0 + \frac{3}{2}v(1) \geq c_0 > 0$. Therefore the above equation gives

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\theta\|^2 + \frac{1}{2}(c_0 - \theta(1))\theta^2(1) &\leq -((u_0 + \delta_0)\rho_x, \theta) \\
&\quad - \frac{3}{4}(v_x\theta, \theta) + \frac{1}{2}(\theta_x\xi, \theta) - \frac{3}{2}(R_{11}, \theta) - \frac{1}{2}(R_{12}, \theta).
\end{aligned} \quad (3.19)$$

Similarly, from (3.18) we obtain

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\xi\|^2 + \frac{1}{2}(-(u_0 - \delta_0 + \frac{3}{2}w(0)) + \xi(0))\xi^2(0) \\
= -((u_0 - \delta_0)\sigma_x, \xi) + \frac{1}{2}(\xi_x\theta, \xi) - \frac{3}{4}(w_x\xi, \xi) - \frac{3}{2}(R_{21}, \xi) - \frac{1}{2}(R_{22}, \xi).
\end{aligned}$$

Again, by (Y2) and (3.5) we get $u_0 - \delta_0 + \frac{3}{2}w(0) \leq -c_0 < 0$. We conclude that

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\xi\|^2 + \frac{1}{2}(c_0 + \xi(0))\xi^2(0) &\leq -((u_0 - \alpha)\sigma_x, \xi) \\
&\quad + \frac{1}{2}(\xi_x\theta, \xi) - \frac{3}{4}(w_x\xi, \xi) - \frac{3}{2}(R_{21}, \xi) - \frac{1}{2}(R_{22}, \xi).
\end{aligned} \quad (3.20)$$

Finally, adding (3.19) and (3.20) we get, as long as the solution of (3.6)-(3.8) exists, that

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} (\|\theta\|^2 + \|\xi\|^2) + \frac{1}{2}(c_0 - \theta(1))\theta^2(1) + \frac{1}{2}(c_0 + \xi(0))\xi^2(0) \\
\leq -((u_0 + \delta_0)\rho_x, \theta) - ((u_0 - \delta_0)\sigma_x, \xi) - \frac{3}{4}(v_x\theta, \theta) - \frac{3}{4}(w_x\xi, \xi) \\
+ \frac{1}{2}(\theta_x\xi, \theta) + \frac{1}{2}(\xi_x\theta, \xi) - \frac{3}{2}(R_{11}, \theta) - \frac{1}{2}(R_{12}, \theta) - \frac{3}{2}(R_{21}, \xi) - \frac{1}{2}(R_{22}, \xi).
\end{aligned} \quad (3.21)$$

In view of (3.8), by continuity we conclude that there exists a maximal temporal instance $t_h > 0$ such that v_h, w_h exist for $t \leq t_h$ and

$$\|\theta(t)\|_{1,\infty} + \|\xi(t)\|_{1,\infty} \leq c_0, \quad t \in [0, t_h]. \quad (3.22)$$

Suppose that $t_h < T$. For $t \in [0, t_h]$ we have by (3.22)

$$\frac{1}{2}(c_0 - \theta(1))\theta^2(1) + \frac{1}{2}(c_0 + \xi(0))\xi^2(0) \geq 0, \quad (3.23)$$

and

$$\frac{1}{2}|(\theta_x \xi, \theta)| + \frac{1}{2}|(\xi_x \theta, \xi)| \leq \frac{c_0}{2} \|\theta\| \|\xi\|. \quad (3.24)$$

We obviously have

$$|(v_x \theta, \theta)| + |(w_x \xi, \xi)| \leq C(\|\theta\|^2 + \|\xi\|^2). \quad (3.25)$$

Using now the approximation and inverse properties (2.1)-(2.6) for \hat{S}_h (and also for \hat{S}_h°) we estimate the rest of the terms in the right-hand side of (3.21) as follows. We first clearly have

$$|((u_0 + \delta_0)\rho_x, \theta)| + |((u_0 - \delta_0)\sigma_x, \xi)| \leq Ch^{r-1}(\|\theta\| + \|\xi\|). \quad (3.26)$$

Integrating by parts we see by (3.15) that

$$\begin{aligned} (R_{11}, \theta) &= ((v\rho)_x, \theta) - ((\rho\theta)_x, \theta) - (\rho\rho_x, \theta) \\ &= v(1)\rho(1)\theta(1) - (v\rho, \theta_x) - \rho(1)\theta^2(1) + (\rho\theta, \theta_x) - (\rho\rho_x, \theta). \end{aligned}$$

Therefore

$$\begin{aligned} |(R_{11}, \theta)| &\leq C\|\rho\|_\infty \|\theta\|_\infty + C\|\rho\|_\infty \|\theta_x\| + \|\rho\|_\infty \|\theta\|_\infty^2 + \|\rho\|_\infty \|\theta\| \|\theta_x\| + \|\rho\|_\infty \|\rho_x\| \|\theta\| \\ &\leq Ch^r \|\theta\|_\infty + Ch^r \|\theta_x\| + Ch^r \|\theta\|_\infty^2 + Ch^r \|\theta\| \|\theta_x\| + Ch^{2r-1} \|\theta\| \\ &\leq Ch^{r-1}(\|\theta\| + \|\theta\|^2). \end{aligned} \quad (3.27)$$

Integration by parts and (3.15) yield for the R_{12} term that

$$(R_{12}, \theta) = (w\rho_x, \theta) - \frac{1}{2}(w_x \theta, \theta) + (v_x \sigma, \theta) + (v_x \xi, \theta) - (\rho_x \sigma, \theta) - (\rho_x \xi, \theta) - (\theta_x \sigma, \theta).$$

Hence, similarly as above

$$\begin{aligned} |(R_{12}, \theta)| &\leq Ch^{r-1} \|\theta\| + C\|\theta\|^2 + Ch^r \|\theta\| + C\|\xi\| \|\theta\| \\ &\quad + Ch^{2r-1} \|\theta\| + Ch^{r-1} \|\xi\|_\infty \|\theta\| + Ch^r \|\theta\|_\infty \|\theta_x\| \\ &\leq Ch^{r-1} \|\theta\| + C\|\theta\|^2 + C\|\xi\| \|\theta\|. \end{aligned} \quad (3.28)$$

Again, using integration by parts and (3.15) for the R_{21} term, we obtain

$$(R_{21}, \xi) = -(w\sigma, \xi_x) - w(0)\sigma(0)\xi(0) + (\sigma\xi, \xi_x) + \sigma(0)\xi^2(0) - (\sigma\sigma_x, \xi).$$

Therefore

$$\begin{aligned} |(R_{21}, \xi)| &\leq C\|\sigma\| \|\xi_x\| + C\|\sigma\|_\infty \|\xi\|_\infty + \|\sigma\|_\infty \|\xi\| \|\xi_x\| + \|\sigma\|_\infty \|\xi\|_\infty^2 + \|\sigma\|_\infty \|\sigma_x\| \|\xi\| \\ &\leq Ch^r \|\xi_x\| + Ch^r \|\xi\|_\infty + Ch^r \|\xi\| \|\xi_x\| + Ch^r \|\xi\|_\infty^2 + Ch^{2r-1} \|\xi\| \\ &\leq Ch^{r-1}(\|\xi\| + \|\xi\|^2). \end{aligned} \quad (3.29)$$

Finally, by (3.15) and integration by parts we have for the R_{22} term

$$(R_{22}, \xi) = (v\sigma_x, \xi) - \frac{1}{2}(v_x \xi, \xi) + (w_x \rho, \xi) + (w_x \theta, \xi) - (\sigma_x \rho, \xi) - (\sigma_x \theta, \xi) - (\rho\xi_x, \xi).$$

Hence,

$$\begin{aligned} |(R_{22}, \xi)| &\leq C\|\sigma_x\| \|\xi\| + C\|\xi\|^2 + C\|\rho\| \|\xi\| + C\|\theta\| \|\xi\| \\ &\quad + \|\sigma_x\| \|\rho\|_\infty \|\xi\| + \|\sigma_x\| \|\theta\|_\infty \|\xi\| + \|\rho\|_\infty \|\xi_x\| \|\xi\| \\ &\leq Ch^{r-1} \|\xi\| + C\|\xi\|^2 + Ch^r \|\xi\| + C\|\theta\| \|\xi\| + Ch^{2r-1} \|\xi\| \\ &\quad + Ch^{r-1} \|\theta\|_\infty \|\xi\| + Ch^r \|\xi_x\| \|\xi\| \\ &\leq Ch^{r-1} \|\xi\| + C\|\xi\|^2 + C\|\theta\| \|\xi\|. \end{aligned} \quad (3.30)$$

By (3.21), taking into account (3.23)-(3.30) we see that

$$\frac{1}{2} \frac{d}{dt} (\|\theta\|^2 + \|\xi\|^2) \leq Ch^{r-1} (\|\theta\| + \|\xi\|) + C(\|\theta\|^2 + \|\xi\|^2), \quad t \in [0, t_h].$$

An application of Gronwall's Lemma and (3.8) yield

$$\|\theta(t)\| + \|\xi(t)\| \leq Ch^{r-1}, \quad t \in [0, t_h], \quad (3.31)$$

from which by inverse assumptions it follows that $\|\theta\|_{1,\infty} + \|\xi\|_{1,\infty} \leq Ch^{r-5/2}$ for $t \in [0, t_h]$. Since it was assumed that $r \geq 3$ this contradicts the maximality of t_h and (3.31) holds for $0 \leq t \leq T$. The estimate (3.9) follows. Since now $\|v - v_h\|_\infty \leq \|\rho\|_\infty + \|\theta\|_\infty \leq Ch^{r-3/2}$ and similarly $\|w - w_h\|_\infty \leq Ch^{r-3/2}$, and since

$$\eta - \eta_h = [\delta_0 + \frac{1}{4}((v - w) + (v_h - w_h))][(v - w) - (v_h - w_h)],$$

we conclude that $\|\eta - \eta_h\| \leq C(\|v - v_h\| + \|w - w_h\|) \leq Ch^{r-1}$. Similarly $\|u - u_h\| \leq \|v - v_h\| + \|w - w_h\| \leq Ch^{r-1}$, and the proof of Proposition 3.1 is now complete. \square

4. NUMERICAL IMPLEMENTATION AND EXPERIMENTS

4.1. Supercritical case. We will implement the standard Galerkin method for the shallow water equations with characteristic boundary conditions in the supercritical case using the space of piecewise linear continuous functions on a uniform mesh in $[0, 1]$ in the usual manner that problems with nonhomogeneous Dirichlet boundary conditions are approximated in practice. For this purpose we let $x_i = ih$, $0 \leq i \leq N$, $Nh = 1$, define $S_h = \{\phi \in C^0[0, 1] : \phi|_{[x_j, x_{j+1}]} \in \mathbb{P}_1, 0 \leq j \leq N-1\}$, and let \mathring{S}_h consist of the functions in S_h that vanish at $x = 0$. We seek $\eta_h, u_h : [0, T] \rightarrow S_h$, the semidiscrete approximation of the solution of (SW1), satisfying for $0 \leq t \leq T$ $\eta_h(0, t) = \eta_0$, $u_h(0, t) = u_0$, and the system of ode's

$$\begin{aligned} (\eta_{ht}, \phi) + (u_{hx}, \phi) + ((\eta_h u_h)_x, \phi) &= 0, \quad \forall \phi \in \mathring{S}_h, \\ (u_{ht}, \phi) + (\eta_{hx}, \phi) + (u_h u_{hx}, \phi) &= 0, \quad \forall \phi \in \mathring{S}_h, \end{aligned} \quad (4.1)$$

with initial values $\eta_h(0) = P\eta^0$, $u_h(0) = Pu^0$, where P is the L^2 projection onto S_h . We discretize this ode ivp in time by the 'classical' explicit 4th-order accurate Runge-Kutta scheme, written in the case of the ode $y' = f(t, y)$, $0 \leq t \leq T$, in the form

$$\begin{aligned} y^{n,1} &= y^n + \frac{k}{2}f(t^n + \frac{k}{2}, y^n), \\ y^{n,2} &= y^n + \frac{k}{2}f(t^n + \frac{k}{2}, y^{n,1}), \\ y^{n,3} &= y^n + kf(t^n + k, y^{n,2}), \\ y^{n+1} &= y^n + k(\frac{1}{6}f(t^n, y^n) + \frac{1}{3}f(t^n + \frac{k}{2}, y^{n,1}) \\ &\quad + \frac{1}{3}f(t^n + \frac{k}{2}, y^{n,2}) + \frac{1}{6}f(t^n + k, y^{n,3})), \end{aligned} \quad (4.2)$$

where k is the time step, $t^n = nk$, $n = 0, 1, \dots, M-1$, $Mk = T$, and y^n approximates $y(t^n)$. Theoretical and numerical evidence from linear stability theory and previous work by the authors, [AD1], [AD2], on similar nonlinear systems, suggests that the resulting fully discrete scheme is stable under a Courant-number restriction of the form $k/h \leq r_0$.

In our first numerical experiment we check the spatial rate of convergence of this fully discrete scheme. We consider (SW1) with $\eta_0 = 1$ and $u_0 = 3$ and add right-hand sides to the pde's so that the exact solution of the ibvp is $\eta(x, t) = xe^{-xt} + \eta_0$, $u(x, t) = (1 - x - \cos(\pi x))e^{2t} + u_0$. With $h = 1/N$, $k = h/10$ (so that the temporal error is negligible), we obtain the L^2 errors and associated rates of convergence of (essentially) the semidiscrete problem at $T = 1$ shown in Table 4.1. The experimental rates of convergence for both

N	η	order	u	order
40	1.243098(-3)	—	5.623510(-3)	—
80	3.110525(-4)	1.99871	1.405648(-3)	2.00024
160	7.778520(-5)	1.99959	3.513979(-4)	2.00006
320	1.944737(-5)	1.99992	8.784876(-5)	2.00001
480	8.643341(-6)	1.99998	3.904381(-5)	2.00001
520	7.364768(-6)	1.99996	3.326806(-5)	2.00001

TABLE 4.1. L^2 errors and spatial orders of convergence, supercritical case.

components of the solution are clearly equal to 2, i.e. superaccurate, as the expected rates for a general quasiuniform mesh would be equal to 1. A numerical study of stability for this example indicates that the errors remain of the same order of magnitude at $T = 1$ up to about $k/h = 0.13$. For larger values of k/h blow-up eventually occurs. It should be noted that for the same test problem the alternative standard Galerkin formulation (2.7)-(2.9) (analyzed in section 2) coupled with the same Runge-Kutta scheme gives L^2 errors that coincide with those of Table 4.1 to at least 5 significant digits; the stability condition was also the same. (Recall that the result of Proposition 2.1 strictly holds for $r \geq 3$ due to the technical requirement in the proof for controlling the $W^{1,\infty}$ norm of the error. The numerical results suggest that the scheme converges for $r = 2$ as well and that the superaccurate order of convergence for a uniform mesh for $r = 2$, proved in [AD2] for an ibvp for the shallow water equations with homogeneous Dirichlet boundary conditions on u , persists in the case of the ibvp (SW1) too.)

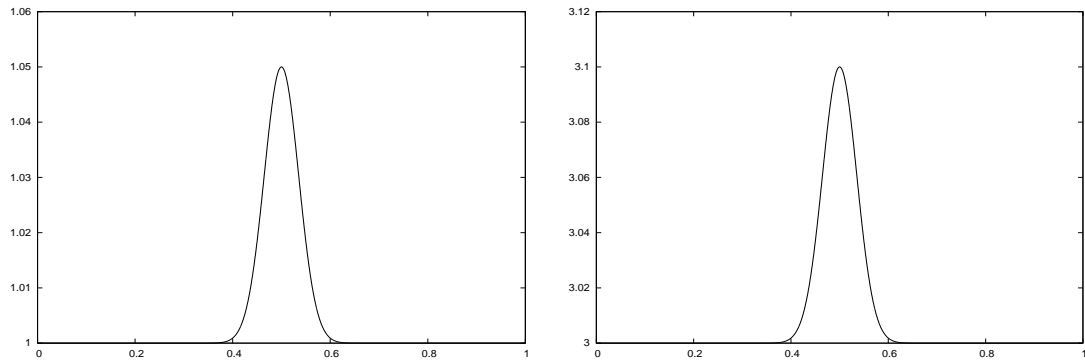
Since the temporal error is much smaller than the spatial one, the experimental estimation of the temporal order of convergence may be done in the following way, used in [BDMK]. Let H_h^n be the fully discrete

k/h	$E^*(T)$	$order$	$E(T)$
1/35	2.6618459890(-8)	—	1.9910684230(-4)
1/40	1.6020860073(-8)	3.8022	1.9910680992(-4)
1/45	1.0112973048(-8)	3.9061	1.9910679532(-4)
1/50	6.6717108025(-9)	3.9478	1.9910678792(-4)
1/55	4.5726218272(-9)	3.9638	1.9910678370(-4)
1/60	3.2362144361(-9)	3.9728	1.9910678102(-4)
1/64	2.5020819256(-9)	3.9865	1.9910677950(-4)
1/64.5	2.4254282105(-9)	3.9983	1.9910677934(-4)
1/65	2.3516195603(-9)	4.0020	1.9910677918(-4)

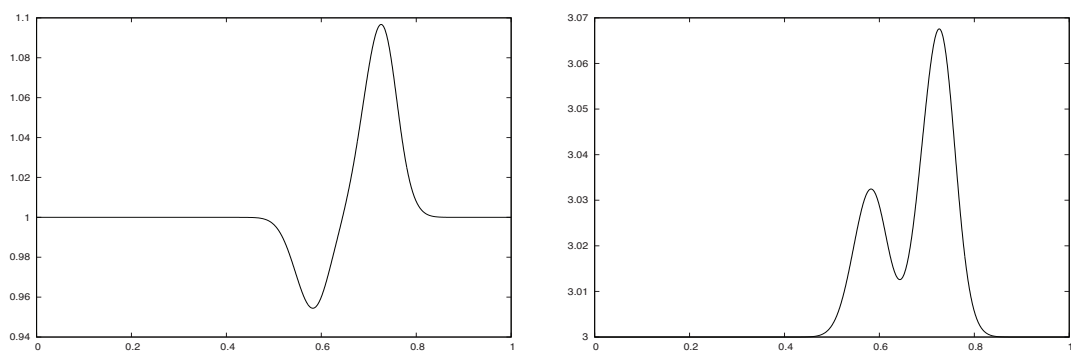
TABLE 4.2. Temporal order of convergence, supercritical case, scheme (4.1)-(4.2), $h = 1/100$, $T = 1$, $k_{ref} = h/120$.

approximation of $\eta(t^n)$. For a fixed value of h we make a reference computation with a very small value $k = k_{ref}$. The approximate solution $H_h^m = H_h^m(h, k_{ref})$, where $mk_{ref} = T$, differs from the exact solution $\eta(\cdot, T)$ by an amount which is practically the error of the spatial discretization. For the same value of h we define a modified L^2 error for small values of k , that are nevertheless considerably larger than k_{ref} , by the formula $E^*(T) = \|H_h^n(h, k) - H_h^m(h, k_{ref})\|$, where $nk = T$. Since taking the difference $H_h^n(h, k) - H_h^m(h, k_{ref})$ essentially cancels the spatial error of $H_h^n(h, k)$, we expect that $E^*(T)$ will decrease at the temporal order of convergence of the scheme as k decreases. This is illustrated in the case of the test problem under consideration and the fully discrete scheme (4.1)-(4.2) in Table 4.2, where $h = 1/100$, $T = 1$, $k_{ref} = h/120$, and $E(T)$ denotes the L^2 error $\|H_h^n(h, k) - \eta(t^n)\|$. For this range of k 's the expected temporal order of convergence, equal to 4, clearly emerges. The analogous experiment with the Galerkin method (2.7)-(2.9) discretized in time with the same Runge-Kutta scheme yields fourth-order temporal convergence in L^2 as well.

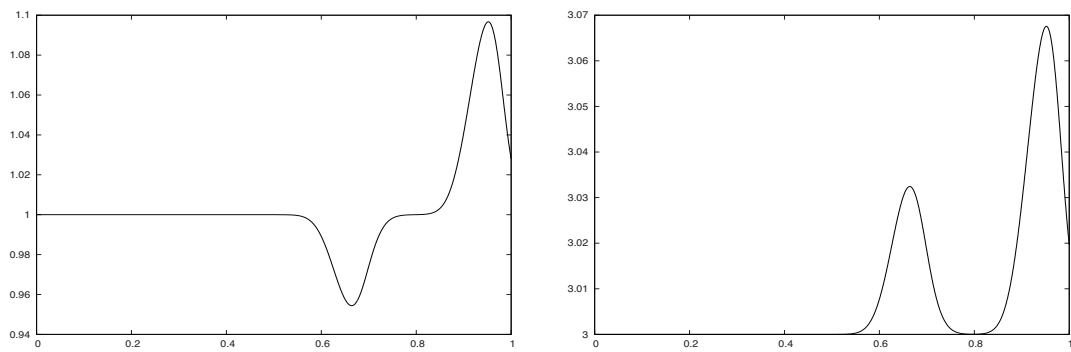
In the next numerical experiment we integrate (SW1) with the fully discrete scheme (4.1)-(4.2), taking $h = 1/N$, $N = 2000$, $k = h/10$, $\eta_0 = 1$, $u_0 = 3$, and initial conditions $\eta^0(x) = 0.05 \exp(-400(x - 0.5)^2) + \eta_0$, $u^0(x) = 0.1 \exp(-400(x - 0.5)^2) + u_0$, $0 \leq x \leq 1$. (Small-amplitude initial conditions were taken to ensure that no discontinuities in the derivatives of the solution develop before the wave profiles exit the spatial interval of integration.) The evolution of the numerical solution is depicted in Figure 4.1(a)-(f). (The approximate η -profiles are on the left and those of u on the right.) The initial Gaussian perturbations of the uniform state $\eta_0 = 1$, $u_0 = 3$ evolve into two unequal pulses for both η and u that travel to the right and exit the computational domain by about $t = 0.4$ without leaving any visible residue or backwards-travelling oscillations as is confirmed by Fig. 4.1(g) that shows the time history of the quantities $\max_x |\eta(x, t) - \eta_0|$ and $\max_x |u(x, t) - u_0|$. (Here η, u denote the approximate solution.) Due to the presence of the spatial and temporal discretizations, the numerical boundary conditions are not expected to be exactly transparent.



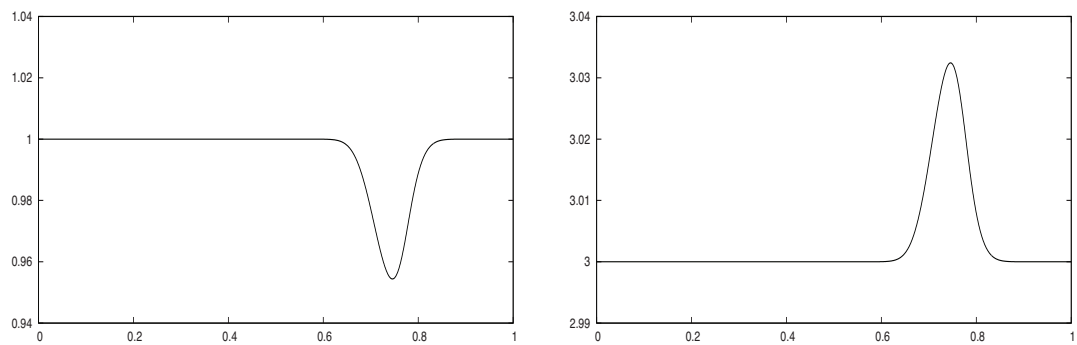
(a) η and u at $t = 0.0$



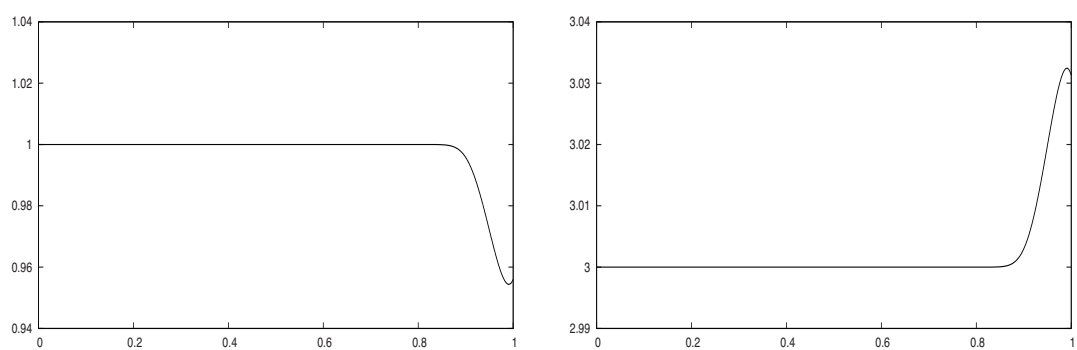
(b) η and u at $t = 0.05$



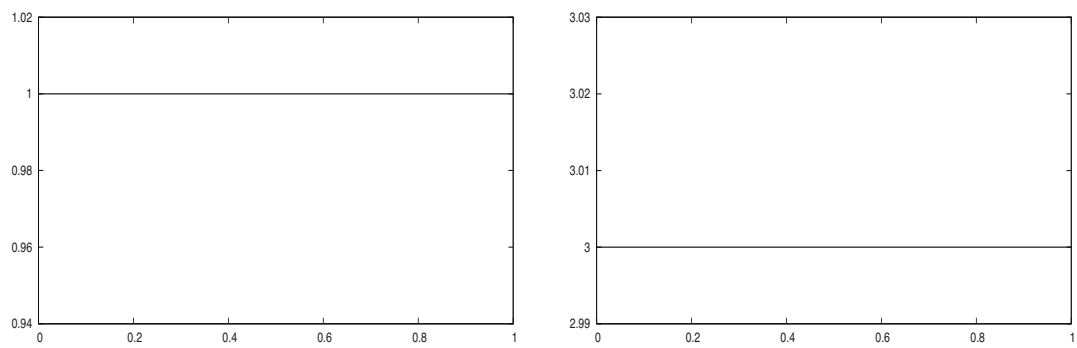
(c) η and u at $t = 0.1$



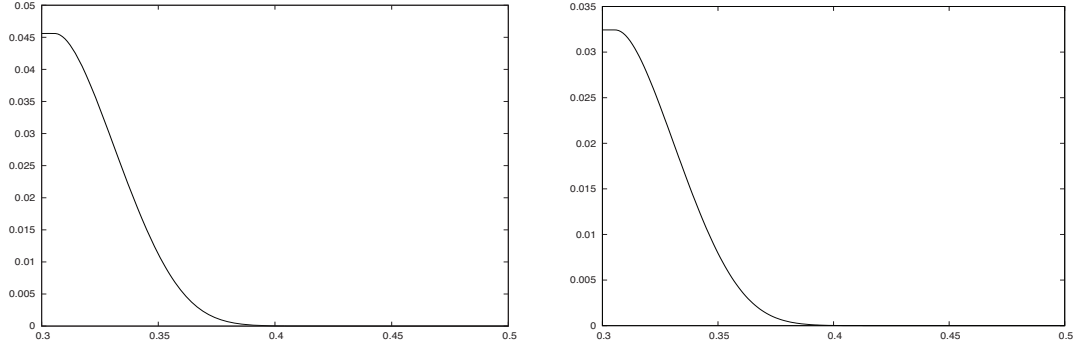
(d) η and u at $t = 0.15$



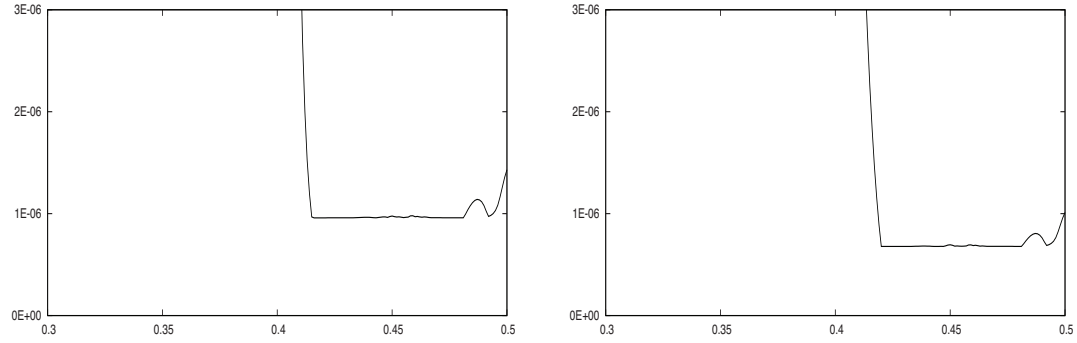
(e) η and u at $t = 0.3$



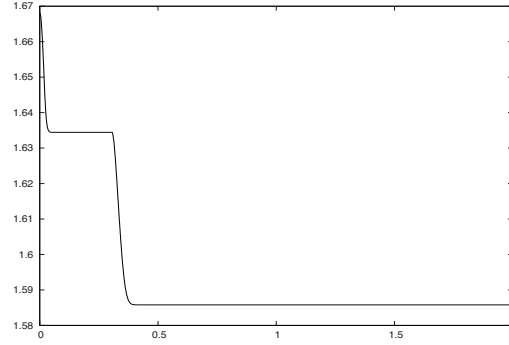
(f) η and u at $t = 0.4$



(g) $\max_x |\eta(x, t) - \eta_0|$ and $\max_x |u(x, t) - u_0|$ vs. time



(h) Magnification of (g)



(i) $\max_x (u - \sqrt{1 + \eta})$ vs. time

FIGURE 4.1. Evolution of Gaussian initial profiles, supercritical case.

However, they are highly absorbing; Figure 4.1(h) reveals that the residue after the waves exit is of $O(10^{-6})$. The positivity of $\max_x (u - \sqrt{1 + \eta})$ for all t checked in Fig.4.1(i) confirms that the numerical solution has remained supercritical throughout the evolution.

In order to study numerically the stability of the fully discrete scheme (4.1)-(4.2) for this test problem, as there are no exact solutions, we took as a measure of error the residual quantity $\max_x |\eta - \eta_0|$. This is plotted in Figure 4.1(g) and (h), stabilizes after the waves exit the computational domain, and has the value

$9.76E - 07$ at $t = 0.45$. We then increased k and observed that this residual was conserved up to about $k/h = 0.3695$ and started increasing afterwards. Since the maximum wave speed c is the speed of the higher rightward-travelling pulse, which is equal to $u + \sqrt{1 + \eta} \simeq 4.5$ for the duration of this experiment, we obtain a Courant number restriction of about $ck/h \leq 1.67$. (Linear stability theory for this method applied to the model problem $\eta_t + c\eta_x = 0$ with periodic boundary conditions and c constant would give a Courant number restriction $ck/h \leq \sqrt{8/3} \simeq 1.633$ which is not far from the experimental result for this small-amplitude nonlinear propagation problem.) It should be noted that the scheme (2.7)-(2.9) discretized in time with the Runge-Kutta method (4.2) yields practically the same numerical results for this test problem.

4.2. Subcritical case. We implement the standard Galerkin method for the SW with characteristic boundary conditions in the subcritical case using again piecewise linear continuous functions on a uniform mesh in $[0, 1]$. In addition to the notation introduced in the previous subsection for the mesh on $[0, 1]$ and the space S_h , we let $S_{h,0}$ consist of the functions in S_h that vanish at $x = 0$ and $x = 1$. We first consider the ‘direct’ standard Galerkin semidiscretization of (SW2), i.e. without reducing first the ibvp into the diagonal form (SW2a). We seek accordingly $\eta_h, u_h : [0, T] \rightarrow S_h$, the semidiscrete approximation of the solution of (SW2), satisfying for $0 \leq t \leq T$

$$(\eta_{ht}, \phi) + (u_{hx}, \phi) + ((\eta_h u_h)_x, \phi) = 0, \quad \forall \phi \in S_h, \quad (4.3)$$

$$(\tilde{u}_{ht}, \chi) + (\eta_{hx}, \chi) + (\tilde{u}_h \tilde{u}_{hx}, \chi) = 0, \quad \forall \chi \in S_{h,0}, \quad (4.4)$$

where $\tilde{u}_h \in S_{h,0}$, and $u_h(x_i, t) = \tilde{u}_h(x_i, t)$, $1 \leq i \leq N - 1$,

$$u_h(x_0, t) = -2\sqrt{1 + \eta_h(x_0, t)} + u_0 + 2\sqrt{1 + \eta_0}, \quad (4.5)$$

$$u_h(x_N, t) = 2\sqrt{1 + \eta_h(x_N, t)} + u_0 - 2\sqrt{1 + \eta_0}. \quad (4.6)$$

At $t = 0$ we compute $u_h(0)$ and $\eta_h(0)$ as the L^2 projections of the initial data η^0, u^0 onto S_h . Although the semidiscrete ivp (4.3)-(4.6) has nonlinear boundary conditions, it is easily discretized in time by an explicit scheme such as the 4th-order RK method (4.2), by first advancing from t^n to t^{n+1} the approximations of η_h and of the ‘interior’ \tilde{u}_h using the temporal discretizations of (4.3) and (4.4), and then updating the $u_h(x_0, t)$ and $u_h(x_N, t)$ values at $t = t^{n+1}$ by (4.5) and (4.6).

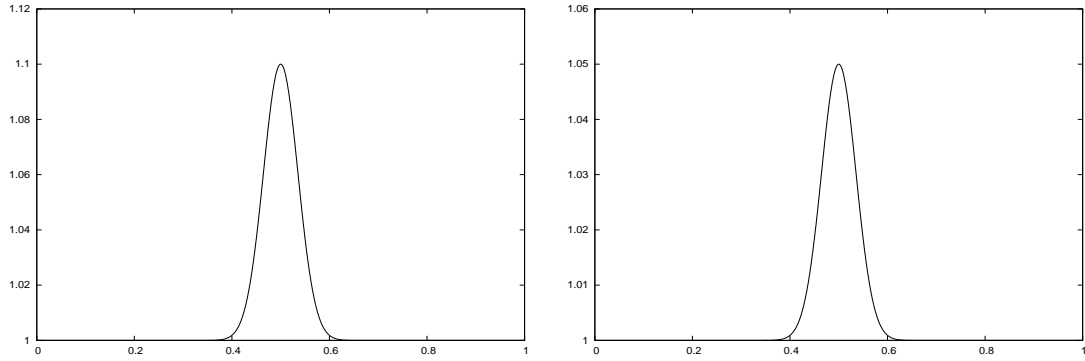
The convergence of this scheme was not analyzed in section 3. The following experiment suggests that in the case of uniform mesh its L^2 errors are of $O(h^2)$. We consider (SW2) with $\eta_0 = 1, u_0 = 1$ and its

N	η	$order$	u	$order$
40	4.847892(-3)	–	2.932354(-3)	–
80	1.207564(-3)	2.00526	7.414336(-4)	1.98367
160	3.017313(-4)	2.00076	1.860285(-4)	1.99479
320	7.544641(-5)	1.99974	4.657627(-5)	1.99786
480	3.353298(-5)	1.99991	2.071174(-5)	1.99867
520	2.857355(-5)	1.99953	1.764866(-5)	1.99944

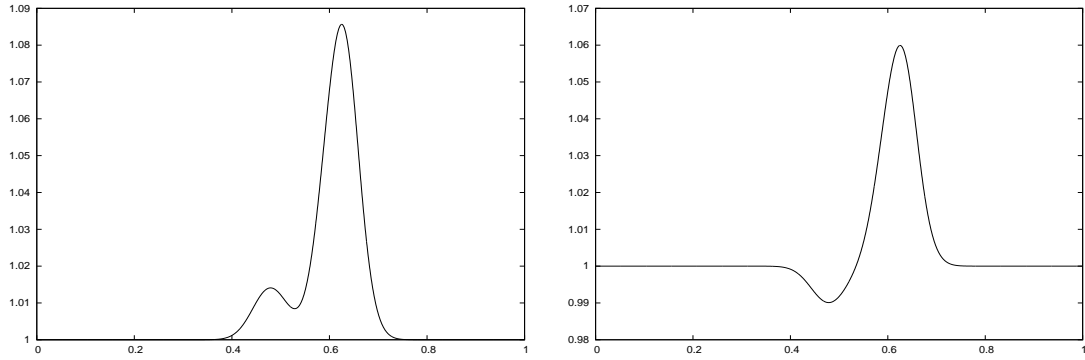
TABLE 4.3. L^2 errors and spatial orders of convergence, subcritical case, semidiscretization (4.3)-(4.6).

semidiscretization (4.3)-(4.6) with piecewise linear continuous functions. We add appropriate right-hand sides to the pde’s in (SW2) so that the exact solution of the ibvp is $\eta(x, t) = (x + 1)e^{-xt}$, $u(x, t) = (2x + \cos(\pi x) - 1)e^t + xA(t) + (1 - x)B(t)$, where $A(t) = 2\sqrt{1 + \eta(1, t)} + u_0 - 2\sqrt{1 + \eta_0}$, $B(t) = -2\sqrt{1 + \eta(0, t)} + u_0 + 2\sqrt{1 + \eta_0}$. We consider uniform spatial and temporal meshes with $h = 1/N$, $k = h/10$, and discretize the semidiscrete problem in time using again the 4th-order ‘classical’ RK scheme. (We checked that the temporal error is negligible for the range of N ’s that we tried.) The resulting L^2 errors and rates of convergence of (essentially) the semidiscrete problem at $T = 1$ are shown in Table 4.3. The rates are practically equal to 2, i.e. superaccurate, as in the supercritical case. An analogous temporal-order calculation to that appearing in Table 4.2 was not so robust and gave rates between 3.7 and 3.9; thus some sort of temporal order reduction cannot be ruled out for this scheme. The experimental Courant number restriction was $k/h \leq 0.53$.

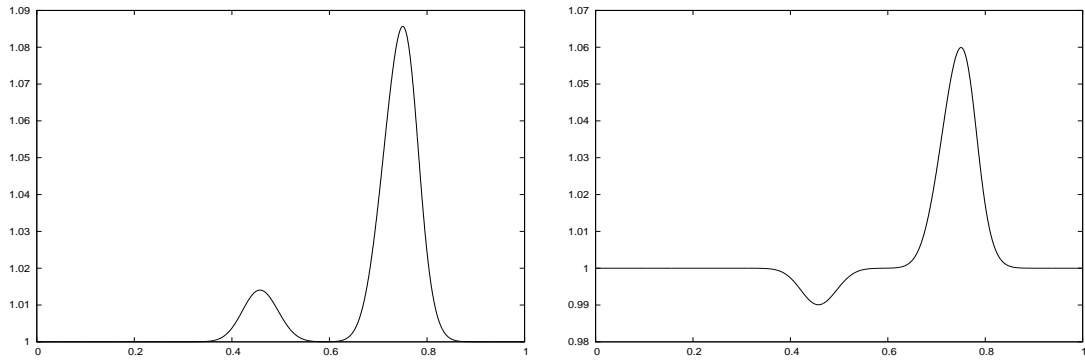
In the following numerical experiment we integrate (SW2) with the same fully discrete scheme, taking $h = 1/N$, $N = 2000$, $k = h/10$, $\eta_0 = u_0 = 1$, and initial conditions $\eta^0(x) = 0.1 \exp(-400(x - 0.5)^2) + \eta_0$, $u^0(x) = 0.05 \exp(-400(x - 0.5)^2) + u_0$. The evolution of the numerical solution is shown in Figure 4.2 (a)-(j).



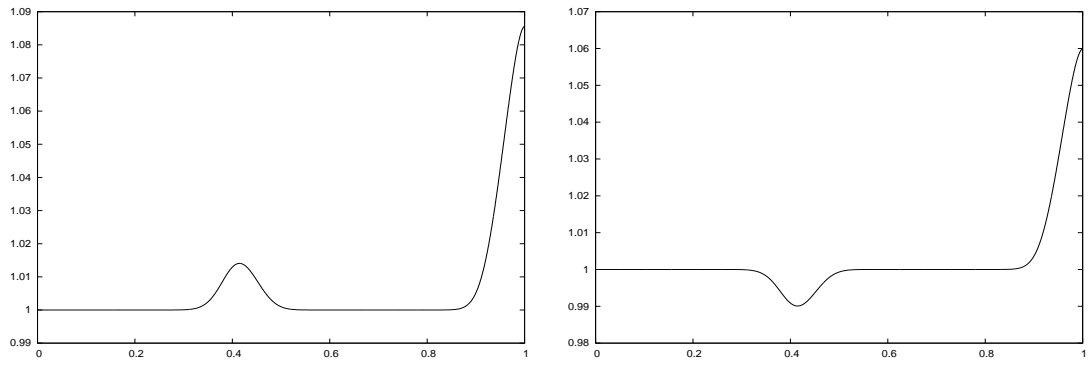
(a) η and u at $t = 0.0$



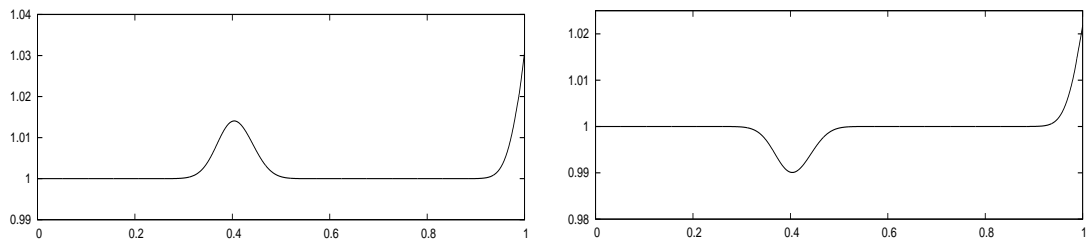
(b) η and u at $t = 0.05$



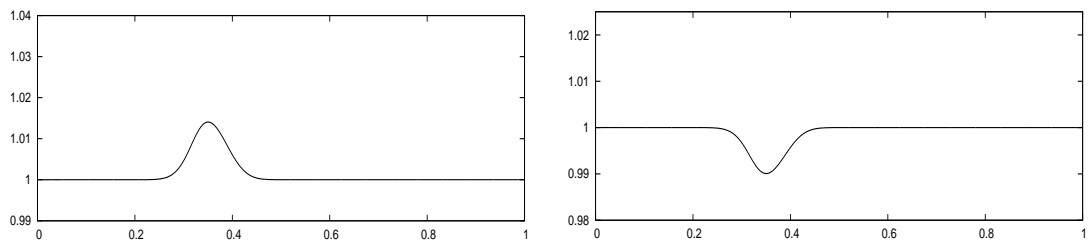
(c) η and u at $t = 0.1$



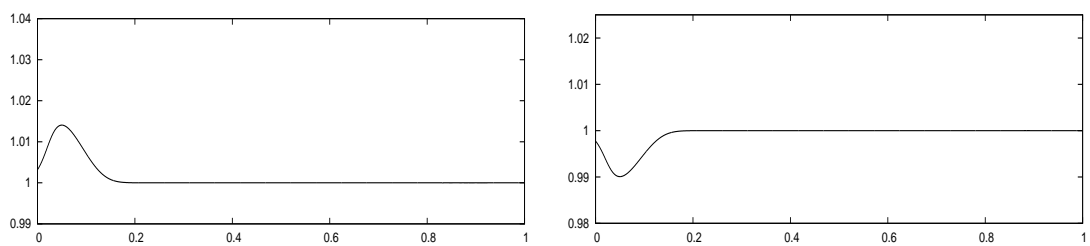
(d) η and u at $t = 0.2$



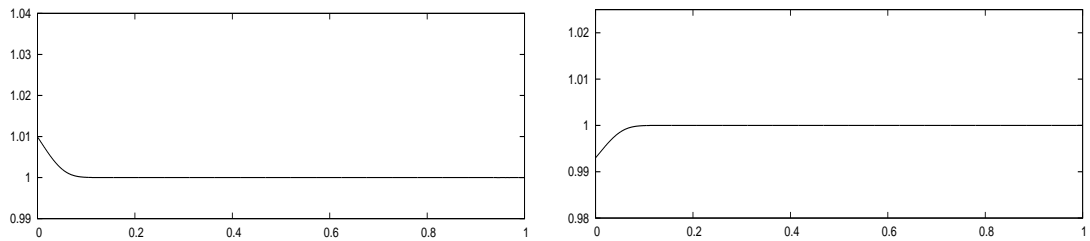
(e) η and u at $t = 0.225$



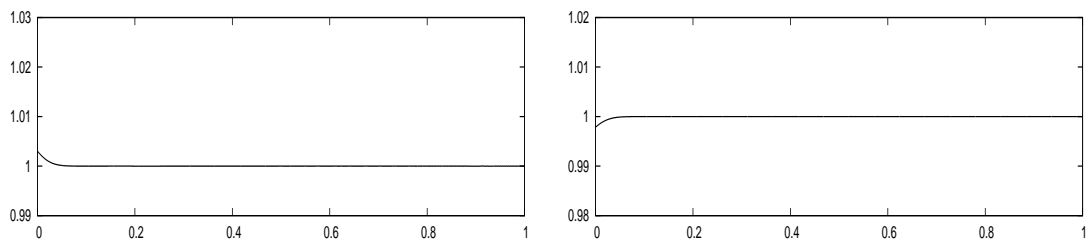
(f) η and u at $t = 0.35$



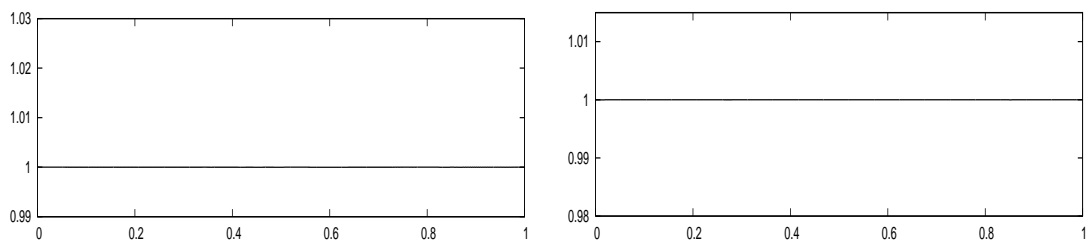
(g) η and u at $t = 1.05$



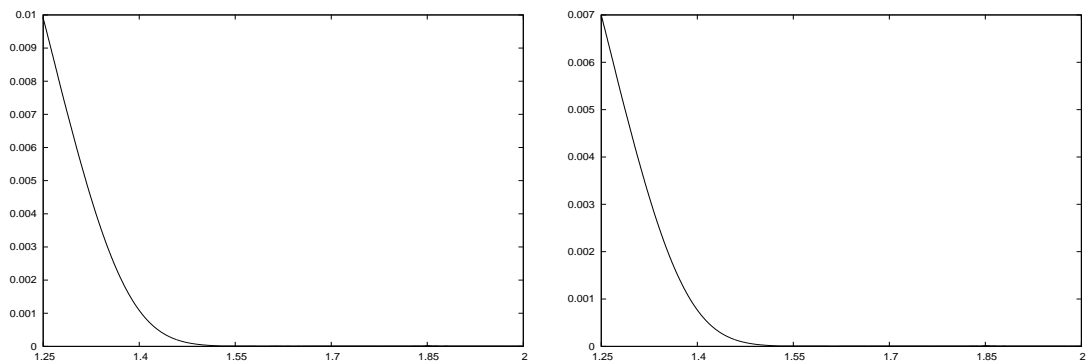
(h) η and u at $t = 1.25$



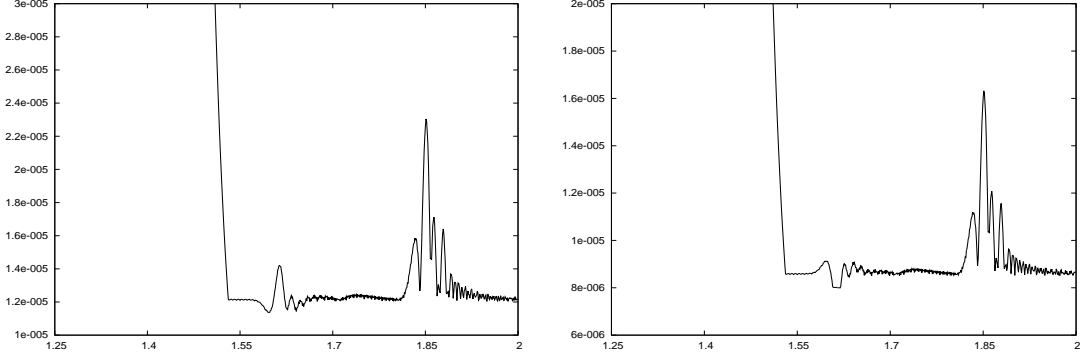
(i) η and u at $t = 1.35$



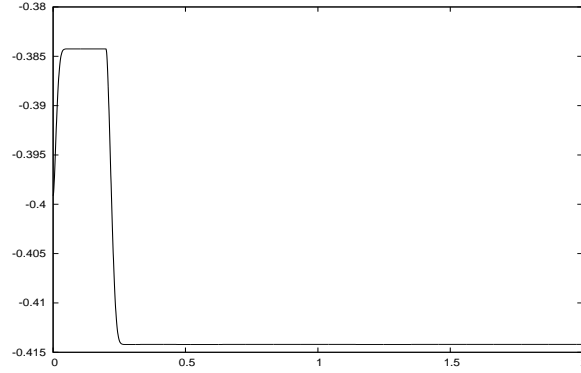
(j) η and u at $t = 1.5$



(k) $\max_x |\eta(x, t) - \eta_0|$ and $\max_x |u(x, t) - u_0|$ vs. time



(l) manification of (k)



(m) $\max_x(u - \sqrt{1 + \eta})$ vs. time

FIGURE 4.2. Evolution of Gaussian initial profiles, subcritical case, semidiscretization (4.3)-(4.6).

The initial Gaussian perturbations of the steady state $\eta_0 = 1$, $u_0 = 3$ evolve into two unequal pulses for both components of the solution, which travel to the right and left and exit the computational domain at about $t = 0.25$ and $t = 1.35$, respectively, without leaving any visible residue as confirmed by the temporal history of the maximum deviations of the approximations of η and u from η_0 and u_0 shown in Figures 4.2(k) and (l). The latter graph shows that the maximum residue after exit is of $O(10^{-5})$, confirming the high degree of absorption of the discrete characteristic boundary conditions. The quantity $\max_x(u - \sqrt{1 + \eta})$ remains negative, i.e. the numerical solution is subcritical, throughout the evolution, cf. Fig. 4.2(m). We investigated the stability of this fully discrete scheme by using again as a measure of error the quantity $\max_x |\eta - \eta_0|$ which stabilizes at $t = 1.55$ to the value $1.21E - 05$. The maximum wave speed c for this problem is about 2.5, and the observed Courant number restriction was $ck/h \leq 1.67$ in conformity with the analogous value in the supercritical case.

In section 3 we analyzed a different Galerkin semidiscretization of the ibvp under consideration, namely that given by (3.6)-(3.8), a semidiscrete approximation of the diagonal form of the ibvp, i.e. of (SW2a). As the following experiment suggests, this semidiscretization is also $O(h^2)$ accurate in L^2 if we use piecewise linear continuous functions on a uniform mesh. We consider the inhomogeneous version of (SW2a) with unknowns $2v$ and $2w$ instead of v and w and take as exact solution the functions $v = u - u_0 + 2(\sqrt{1 + \eta} - \delta_0)$, $w = u - u_0 - 2(\sqrt{1 + \eta} - \delta_0)$, where $\delta_0 = \sqrt{1 + \eta_0}$ and $\eta(x, t) = (x + 1)e^{-xt}$, $u(x, t) = (2x + \cos(\pi x) - 1)e^t + xA(t) + (1 - x)B(t)$, with $A(t) = 2\sqrt{1 + \eta(1, t)} + u_0 - 2\delta_0$, $B(t) = -2\sqrt{1 + \eta(0, t)} + u_0 + 2\delta_0$. For $\eta_0 = u_0 = 1$ we approximate this ibvp by the nonhomogeneous analog of the semidiscretization (3.6)-(3.8) (with unknowns $2v_h$ and $2w_h$) using piecewise linear continuous elements with $h = 1/N$, and the ‘classical’ 4th-order RK scheme for time stepping with $k = h/10$. The resulting L^2 errors and rates of convergence at $T = 1$ are given in Table 4.4. The rates are practically equal to 2 as in the previous cases, due to the

N	η	order	u	order
40	2.470369(−3)	—	9.918820(−4)	—
80	6.172661(−4)	2.00076	2.472869(−4)	2.00398
160	1.543038(−4)	2.00012	6.179903(−5)	2.00053
320	3.857665(−5)	1.99997	1.545737(−5)	1.99929
480	1.714531(−5)	1.99998	6.870865(−6)	1.99967
520	1.460903(−5)	1.99999	5.854663(−6)	1.99958

TABLE 4.4. L^2 errors and spatial orders of convergence, subcritical case, semidiscretization (3.6)-(3.8).

uniform mesh. The temporal order of convergence for this scheme was found to be practically equal to 4; the associated temporal-order calculation with $h = 1/50$, $k_{ref} = h/200$ was quite robust.

We now compare the solutions of the two semidiscretizations (4.3)-(4.6) and (3.6)-(3.8) by means of a numerical experiment. We consider the ibvp (SW2) with $\eta_0 = u_0 = 1$ and initial values $\eta(x, 0) = 0.1 \exp(-400(x - 0.5)^2) + \eta_0$, $u(x, 0) = 0.05 \exp(-400(x - 0.5)^2) + u_0$. (Recall that the temporal evolution of the numerical solution of this ibvp generated by (4.3)-(4.6) with $N = 2000$, $h = 1/N$, $k = h/10$ is shown in Figure 4.2.) We will compare the numerical solutions of this ibvp with both discretizations, using the same numerical parameters. Let (η_h, u_h) denote the fully discrete approximations at time t produced by (4.3)-(4.6) as underlying semidiscretization. In addition, let (η_{hD}, u_{hD}) be functions in S_h with point values computed by the formulas (3.10), where v_h, w_h are now the fully discrete approximations at t produced when we use the method (3.6)-(3.8). Define $\varepsilon(t) = \max_{0 \leq i \leq N} |\eta_h(x_i, t) - \eta_{hD}(x_i, t)|$ and $e(t) = \max_{0 \leq i \leq N} |u_h(x_i, t) -$

t	0.0	0.05	0.1	0.15	0.2	0.225	0.25
$\varepsilon(t)$	2.0(−8)	2.0(−8)	2.0(−8)	2.0(−8)	4.0(−8)	3.0(−8)	3.0(−8)
$e(t)$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	2.0(−8)	2.0(−8)	2.0(−8)

t	0.35	0.75	1.05	1.15	1.25	1.35	1.5
$\varepsilon(t)$	1.232(−5)	1.301(−5)	1.237(−5)	1.3(−5)	1.224(−5)	1.23(−5)	1.217(−5)
$e(t)$	8.71(−6)	9.2(−6)	8.75(−6)	9.54(−6)	8.65(−6)	8.75(−6)	8.6(−6)

TABLE 4.5. Comparison of the discretizations (4.3)-(4.6) and (3.6)-(3.8), subcritical case, experiment of Fig. 4.2

$u_{hD}(x_i, t)|$. The quantities ε and e for various values of $t = t^n \in [0, 1]$ are given in Table 4.5. We observe that up to about $t = 0.25$ the errors are of $O(10^{-8})$ and subsequently increase to $O(10^{-5})$. This is due to the fact that about $t = 0.25$ the larger, rightwards-travelling wave completes its exit from the computational domain (see Figure 4.2(e), and we expect a small residue to be radiated into $[0, 1]$ because the numerical characteristic boundary conditions of (4.3)-(4.6) are not exactly transparent. This residue has a magnitude of $O(10^{-5})$ as evidenced by Fig. 4.2(l). It is worthwhile to note that the ‘diagonal’ approximation (3.6)-(3.8) leaves a practically negligible residue. This is suggested by the evidence in Figure 4.3. In this figure, the two graphs on the left depict what is left in the computational domain of the η - and u -components of the numerical solution generated by (4.3)-(4.6) at $t = 1.5$, after the waves have exited the domain, while those on the right are the analogous η - and u -profiles generated by (3.6)-(3.8). (Thus the graphs on the left are magnifications of the graphs of Fig. 4.2(j).) We observe that the residues of the usual Galerkin semidiscretizations are of $O(10^{-5})$ while those of the ‘diagonal’ scheme are much smaller. It is clear, at least for this example, that the ‘diagonal’ Galerkin method has practically transparent boundary conditions.

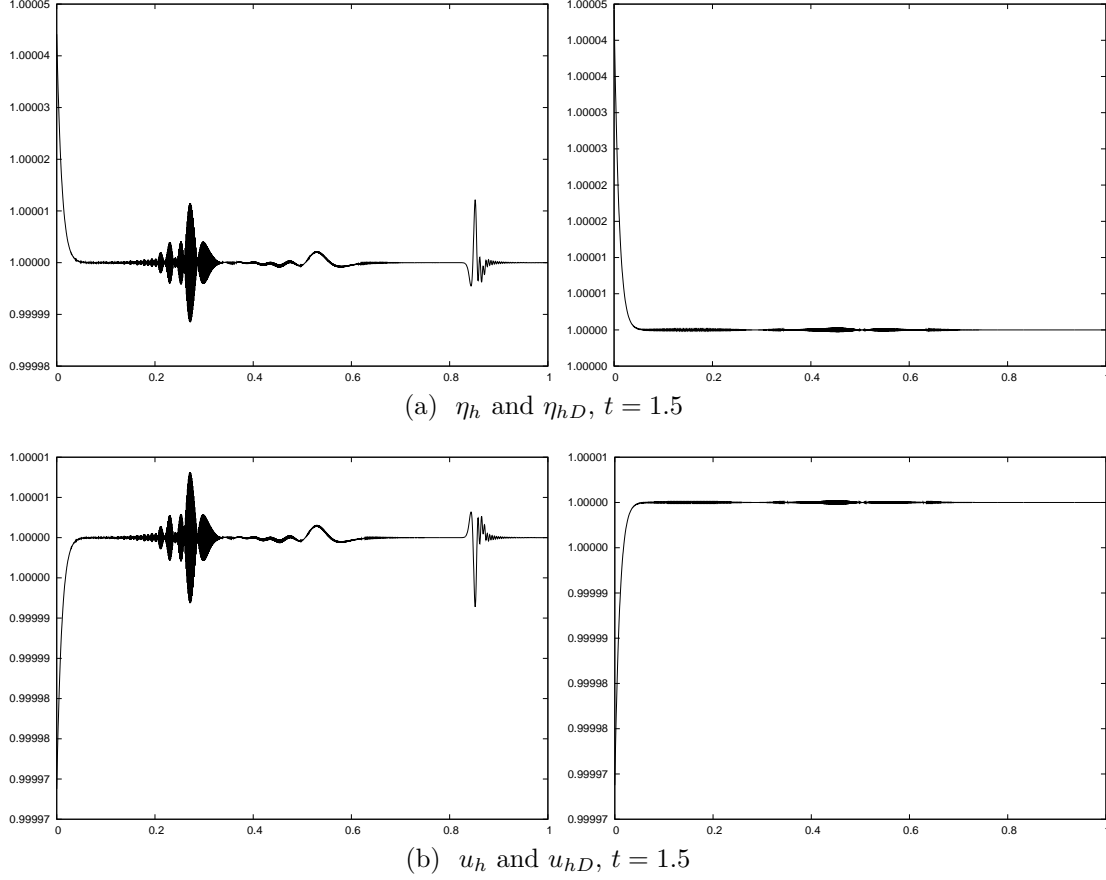


FIGURE 4.3. Magnifications of the profiles of the numerical solutions generated by the methods (4.3)-(4.6) (left) and (3.6)-(3.8) (right) at $t = 1.5$. Subcritical case, evolution as in Fig. 4.2.

4.3. Linearized vs. nonlinear characteristic boundary conditions in the subcritical case. If the elevation of the free surface η is a small perturbation of the steady state η_0 one may derive *linearized* approximations to the characteristic boundary conditions: Assume that the wave height is given by $1 + \eta = 1 + \eta_0 + \tilde{\eta}$, where $|\tilde{\eta}| \ll 1 + \eta_0$. Then $\sqrt{1 + \eta} = \sqrt{1 + \eta_0} + \frac{\eta - \eta_0}{2\sqrt{1 + \eta_0}} + O(\tilde{\eta}^2)$, and, in the subcritical case (SW2), from the characteristic boundary condition, for example at $x = 1$, it follows that

$$u(1, t) - \frac{\eta(1, t)}{\sqrt{1 + \eta_0}} = u_0 - \frac{\eta_0}{\sqrt{1 + \eta_0}} + O(\tilde{\eta}^2),$$

from which neglecting the $O(\tilde{\eta}^2)$ term we obtain the linearized b.c.

$$u(1, t) - \frac{\eta(1, t)}{\sqrt{1 + \eta_0}} = u_0 - \frac{\eta_0}{\sqrt{1 + \eta_0}}, \quad (4.7)$$

and similarly at $x = 0$ the b.c.

$$u(0, t) + \frac{\eta(0, t)}{\sqrt{1 + \eta_0}} = u_0 + \frac{\eta_0}{\sqrt{1 + \eta_0}}. \quad (4.8)$$

The linearized boundary conditions (4.7) and (4.8) have often been used in the computational fluid dynamics literature, as was mentioned in the Introduction. It is not hard to see that (4.7) and (4.8) are transparent for the *linearized* shallow water equations obtained by linearizing the system (1.1) about the steady state (u_0, η_0) . This may be seen by diagonalizing the linearized system and explicitly computing its solutions propagating along the characteristics. In [SLTT] it is proved that the ‘energy’ integral $\int_0^1 (u^2 +$

$\frac{1}{1+\eta_0}\eta^2)dx$ of the solution (η, u) of the linearized system decreases with time in the presence of a class of boundary conditions that includes (4.7)-(4.8), a fact that implies the well-posedness of the ibvp for the linearized system supplemented by (4.7) and (4.8).

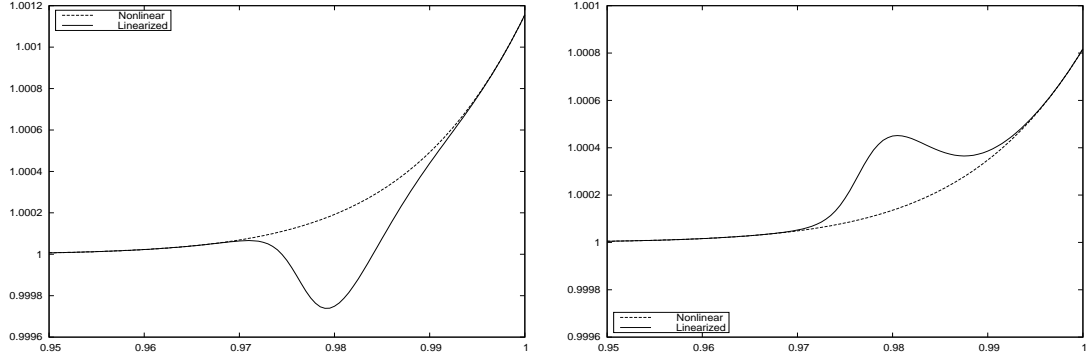
However, the linearized characteristic boundary conditions are not transparent for the nonlinear system (1.1). So, there arises a need to study their absorption properties. In [NMF] and [SLTT] the linearized conditions were compared with the nonlinear ones, appropriately discretized in the finite difference schemes used in these two references. The comparison was effected by means of numerical experiments, many of which in more complicated instances of hydraulic and geophysical interest, including single- and multi- layered flows in the presence of variable bottom topography, cross velocity terms, and Coriolis forces. Moreover, the finite difference schemes used in [NMF] and [SLTT] allow in general the simulation of flows that develop steep fronts and discontinuities. It was confirmed that the linearized conditions in many examples give rise to spurious oscillations that are reflected backwards into the computational domain.

In the sequel we will compare the two sets of boundary conditions in the case of smooth subcritical flows discretized by the ‘direct’ fully discrete Galerkin method with piecewise linear continuous functions in space coupled with ‘classical’ 4th-order RK time stepping. The spatial discretization in the nonlinear b.c. case is given by (4.3)-(4.6); in the linearized case (4.5)-(4.6) are replaced by their obvious linearized analogs resulting from (4.7)-(4.8). We first check the L^2 errors and associated orders of convergence for the scheme with the linearized b.c.’s. Adding appropriate right-hand sides to the pde’s in (SW2) so that the exact solution of the ibvp is $\eta(x, t) = (x + 1)e^{-xt}$, $u(x, t) = (2x + \cos(\pi x) - 1)e^t + xa(t) + (1 - x)b(t)$, where $a(t) = u_0 + \frac{2e^{-t} - \eta_0}{\sqrt{1 + \eta_0}}$, $b(t) = u_0 + \frac{\eta_0 - 1}{\sqrt{1 + \eta_0}}$, we take $u_0 = \eta_0 = 1$ and compute with $h = 1/N$ and $k = h/20$. The resulting L^2 errors of (essentially) the spatial discretization are shown, accompanied by the associated orders of convergence in Table 4.6. Due to the spatial uniform mesh the rates are again practically equal to 2.

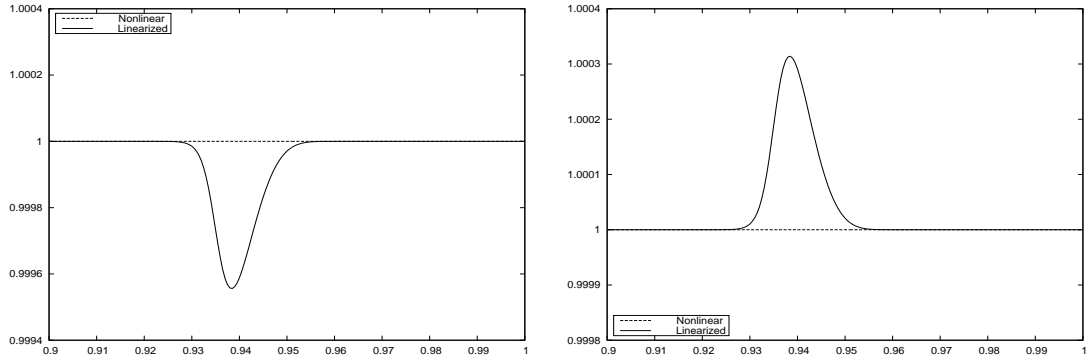
N	η	$order$	u	$order$
40	4.835002(−3)		2.930984(−3)	
80	1.204245(−3)	2.00539	7.408500(−4)	1.98413
120	5.350289(−4)	2.00088	3.299726(−4)	1.99472
160	3.008683(−4)	2.00099	1.858783(−4)	1.99497
200	1.925597(−4)	1.99991	1.190431(−4)	1.99695
240	1.337304(−4)	1.99966	8.269369(−5)	1.99835
280	9.825114(−5)	1.99998	6.077487(−5)	1.99783
320	7.523153(−5)	1.99920	4.653473(−5)	1.99936
360	5.944094(−5)	2.00018	3.677599(−5)	1.99820
400	4.814899(−5)	1.99964	2.979144(−5)	1.99908
440	3.979234(−5)	2.00006	2.462384(−5)	1.99880
480	3.343735(−5)	1.99975	2.069270(−5)	1.99898
520	2.849223(−5)	1.99946	1.763197(−5)	1.99978

TABLE 4.6. L^2 errors and spatial orders of convergence, subcritical case, semidiscretization (4.3), (4.4) with the linearized b.c. (4.7), (4.8).

We now consider a numerical experiment discussed in Section 4.2, wherein the fully discrete numerical scheme based on the semidiscretization (4.3)-(4.6) gives the evolution depicted in Figure 4.2. We repeat the experiment with the linearized boundary conditions (4.7), (4.8) instead of (4.5),(4.6). In Figure 4.4 the graphs of η_{lin} , u_{lin} (solid line), i.e. the numerical solution corresponding to the linearized b.c.’s, are superimposed on those of the numerical solution η_{Nlin} , u_{Nlin} (dotted line) computed with the nonlinear b.c.’s. At $t = 0.25$, when the rightwards-travelling pulse has almost completed its exit from the computational domain at $x = 1$ (recall Fig. 4.2(g)), we observe that the linearized b.c.’s give rise to a spurious pulse of amplitude about 4×10^{-4} that is reflected inwards and travels (as Fig. 4.4(b) confirms) to the left with a speed of $-u_0 + \sqrt{1 + \eta_0} \cong 0.41$ as predicted by the analysis of the linearized system of which this small-amplitude pulse is an approximate solution. A similar (but much smaller) rightwards-travelling pulse is created at $x = 0$ when the smaller wave in Fig. 4.2 exits the computational interval. (These pulses will



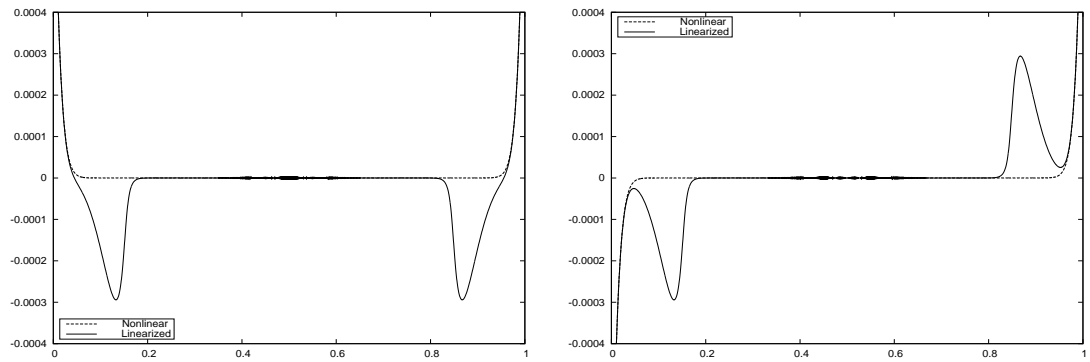
(a) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 0.25$



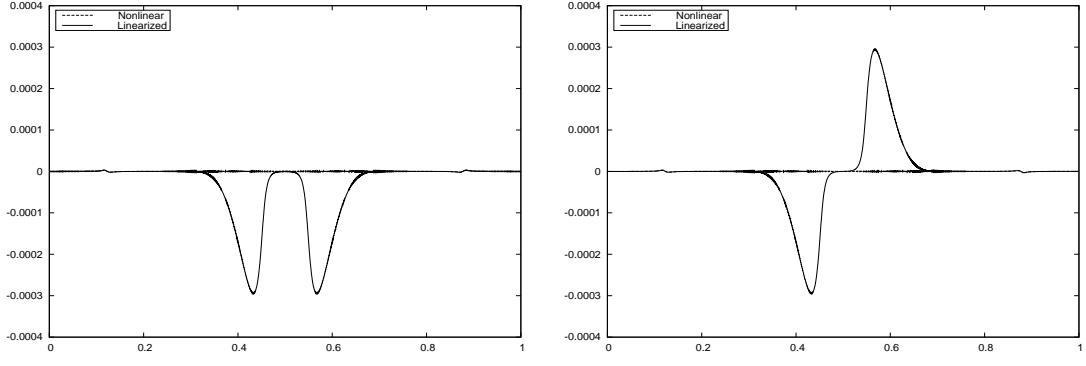
(b) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 0.35$

FIGURE 4.4. Reflected spurious pulses due to the linearized b.c.'s (solid lines) superimposed on the solution of (4.3)-(4.6) (dotted lines); evolution of Fig. 4.2.

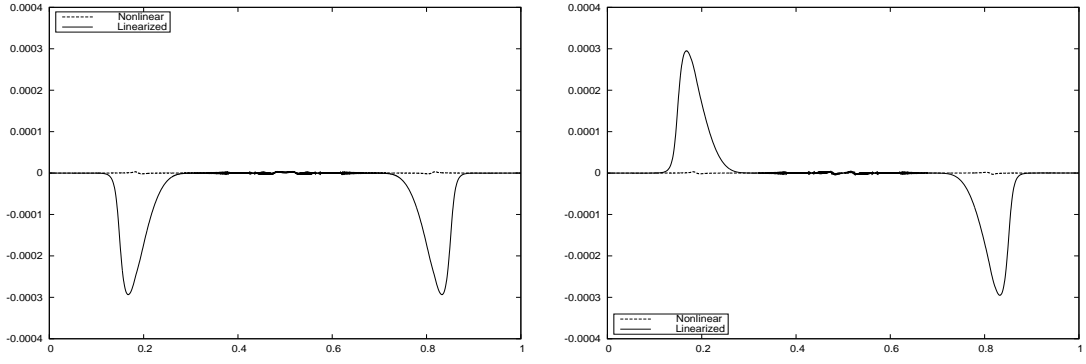
eventually exit the interval since the linearized b.c.'s are transparent for the linearized system.) This may also be seen in the similar numerical experiment shown in Figure 4.5. We consider (SW2) with $\eta_0 = u_0 = 0$



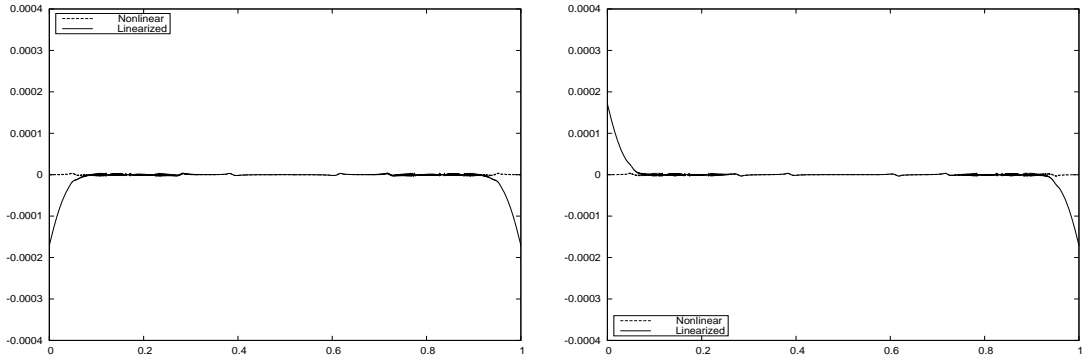
(a) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 0.6$



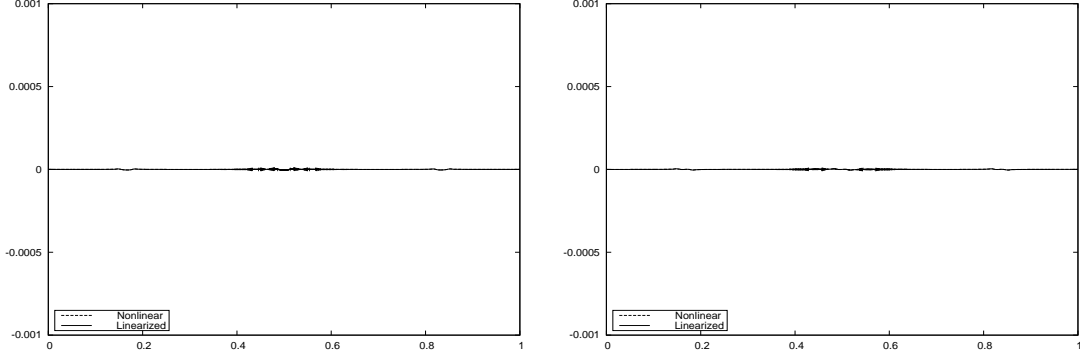
(b) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 0.9$



(c) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 1.3$



(d) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 1.5$



(e) $\eta_{lin.}$ vs. $\eta_{Nlin.}$ and $u_{lin.}$ vs. $u_{Nlin.}$ at $t = 1.6$

FIGURE 4.5. Reflected spurious pulses due to the linearized b.c.'s (solid lines) superimposed on the solution of (4.3)-(4.6) (dotted lines); evolution resulting from $\eta^0 = 0.1 \exp(-400(x - 0.5)^2)$, $u^0 = 0$, $\eta_0 = u_0 = 0$.

and initial conditions $\eta^0(x) = 0.1 \exp(-400(x - 0.5)^2)$, $u^0(x) = 0$, that we solve by the fully discrete Galerkin method corresponding to (4.3)-(4.4) with nonlinear and linearized b.c.'s using $h = 1/N$, $k = h/10$, $N = 2000$. Due to symmetry the initial Gaussian $\eta^0(x)$ breaks up into two waves that travel with equal speeds in opposite directions and exit the computational domain at about $t = 0.6$. Figure 4.5 show magnifications of the ensuing behavior of the numerical solution computed with the linearized b.c.'s (solid line) superimposed on the correct steady state (dotted line), the result of computing with the nonlinear b.c.'s. Two equal spurious pulses are reflected inwards at both ends and, according to linear theory, travel with unit speed, interact linearly, and exit cleanly by $t = 1.6$.

We consider next the shallow water equations written in dimensional variables (h, u) where u is the horizontal velocity and h denotes now the height of the water column above the bottom; the latter is located at a depth H below the level of rest. The system, posed on a channel of length $2L$, is written, in the notation of [NMF], in the form

$$\begin{aligned} h_t + (hu)_x &= 0, \\ u_t + gh_x + uu_x &= 0, \end{aligned} \quad -L \leq x \leq L, \quad t \geq 0, \quad (4.9)$$

with initial conditions

$$h(x, 0) = f(x), \quad u(x, 0) = v(x), \quad -L \leq x \leq L, \quad (4.10)$$

where g is the acceleration of gravity. The system is supplemented by the nonlinear characteristic boundary conditions now written as

$$u(-L, t) + 2\sqrt{gh(-L, t)} = a_E^+, \quad (4.11)$$

$$u(L, t) - 2\sqrt{gh(L, t)} = a_E^-, \quad (4.12)$$

where $a_E^\pm = u_0 \pm 2\sqrt{gh_0}$, and u_0, h_0 are the constant values of u and h outside $[-L, L]$, i.e. the 'asymptotic' state of the flow. The linearized boundary conditions (obtained by linearizing (4.11)-(4.12) assuming that $h = H + \tilde{h}$, where \tilde{h} is small) are

$$u(-L, t) + \sqrt{\frac{g}{H}}h(-L, t) = b_E^+, \quad (4.13)$$

$$u(L, t) - \sqrt{\frac{g}{H}}h(L, t) = b_E^-, \quad (4.14)$$

where $b_E^\pm = u_0 \pm \sqrt{\frac{g}{H}}h_0$. We repeat the numerical experiment in section 4.1 of [NMF], using now the fully discrete Galerkin method with piecewise linear continuous functions in space and the 4th-order classical RK method in time, and taking $L = 2$ m, $H = 0.2$ m, $g = 9.8$ m/sec², $h_0 = 0.2$ m, $u_0 = 0$, i.e. $a_E^\pm = \pm 2.8$ m/sec, $b_E^\pm = \pm 1.4$ m/sec, and initial values $f(x) = 0.25$ m and $v(x) = 0$, with discretization parameters $\Delta x = 4/N$,

$k = \Delta t = 1/(10N)$, $N = 8000$. We first solve the problem with the nonlinear b.c.'s (4.11), (4.12). The evolution of the waveheight h is depicted in Fig. 4.6. Due to the difference of the initial profile $h(x, 0) = f(x)$

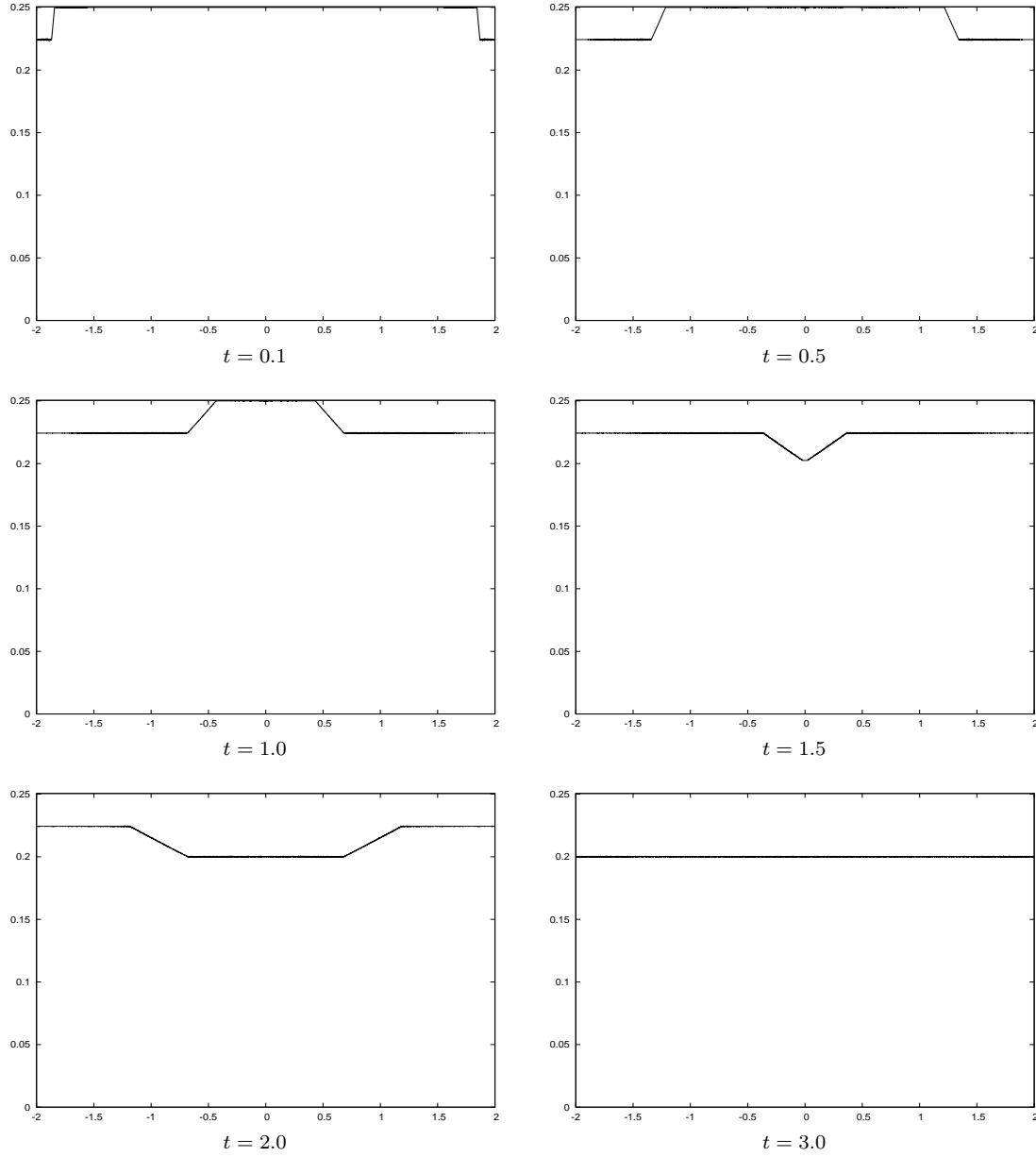


FIGURE 4.6. Evolution of waveheight $h(x, t)$; numerical solution of (4.9)-(4.12) with $f(x) = 0.25$ m, $v(x) = 0$, $h_0 = 0.2$ m, $u_0 = 0$.

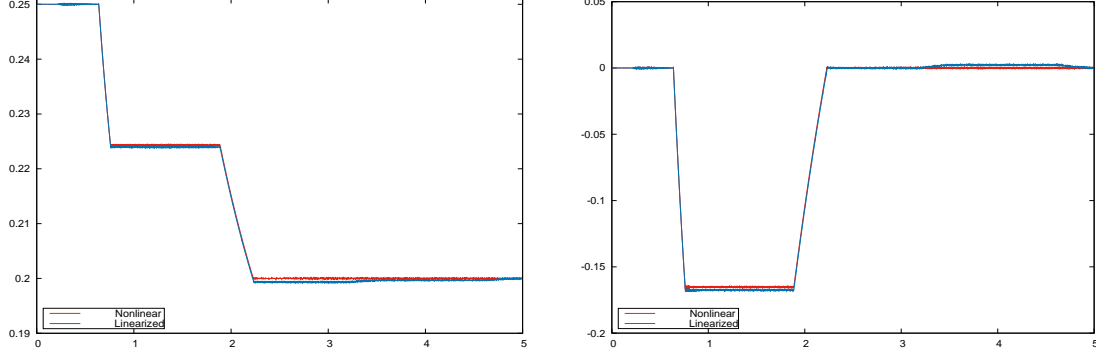
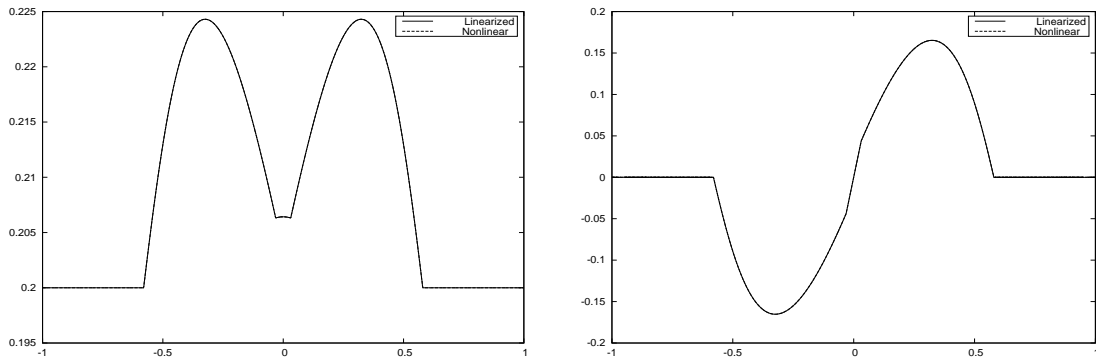


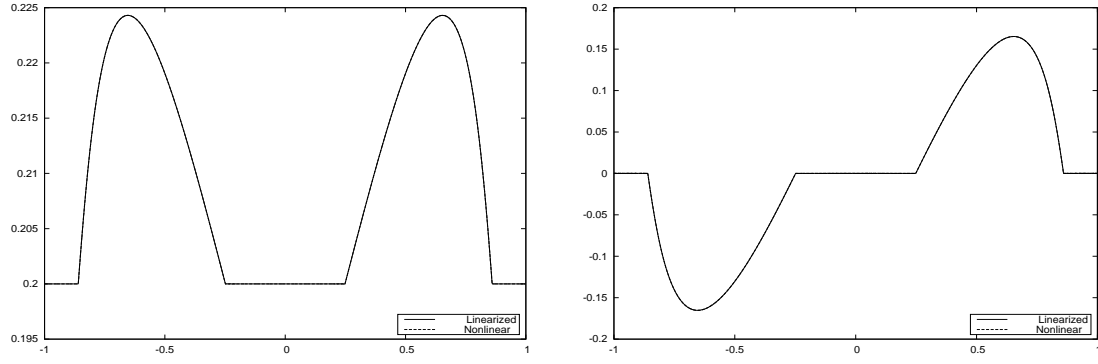
FIGURE 4.7. Graphs of $h(-1, t)$ and $u(-1, t)$ vs. time. Evolution of Figure 4.6, nonlinear and linearized characteristic b.c.'s

at $x = \pm 2$ with the asymptotic state h_0 , fluid exits the domain from both boundary points and two wavefronts are created that travel in opposite directions, interact, and completely exit the computational domain at about $t = 2.25$ sec with no apparent reflections. After this time the asymptotic state $h_0 = 0.2$ m is achieved throughout the domain. The time history of the solution (h, u) at $x = -1$ is given in Figure 4.7 and corresponds to the three stages of the evolution deduced from Figure 4.6. The graph of the numerical solution corresponding to the linearized b.c.'s (4.13)–(4.14) is superimposed on the previous graph. It may be seen that solution with the linearized b.c.'s is just slightly less absorbing. These results are in agreement with those of [NMF].

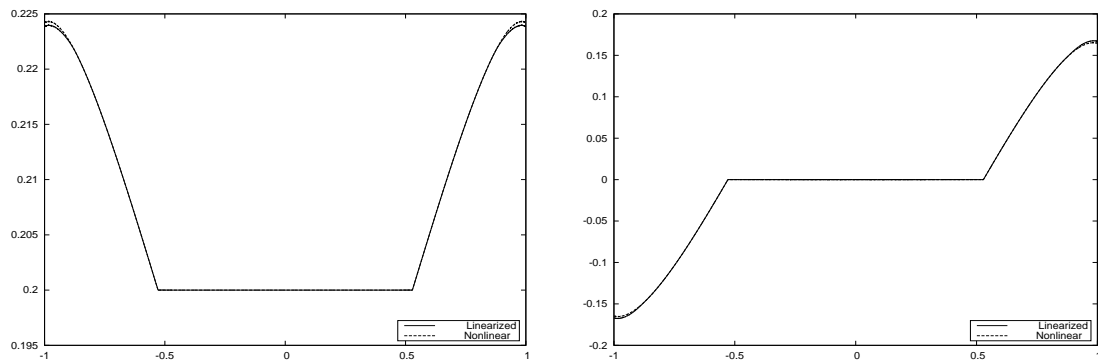
In our final numerical experiment we solve again the SW on $[-L, L]$ in their dimensional form (4.9), (4.10) with the nonlinear b.c.'s (4.11), (4.12) and the linearized ones (4.13), (4.14), taking now $L = 1$ m, $H = 0.2$ m, $g = 9.8$ m/sec², $h_0 = 0.2$ m, $u_0 = 0$, i.e. $a_E^\pm = \pm 2.8$ m/sec, $b_E^\pm = \pm 1.4$ m/sec, and as initial values in (4.10) $v(x) = 0$ and $f(x)$ given by a half-sine pulse centered at $x = 0$ and equal to $0.2 + 0.05 \sin(\pi(x + 0.3)/0.6)$ m if $|x| \leq 0.3$ m and to 0.2 m for $|x| > 0.3$ m. The mesh parameters are $\Delta x = 2/N$, $N = 4000$ and $\Delta t = 1/(10N)$. Figure 4.8 shows the ensuing evolution of the numerical solution (h is shown on the left and u on the right at each temporal instance.) The graphs corresponding to the nonlinear characteristic b.c.'s (dotted line) are superimposed on those computed with the linearized b.c.'s. Two pulses are produced (with no oscillations apparent at the points where the spatial derivative is discontinuous) that travel to opposite directions and exit the computational interval $[-1, 1]$ with no reflections in the case of the nonlinear b.c.'s, shortly after $t = 0.8$ sec. The linearized b.c.'s give spurious reflected pulses with waveheights of amplitude of $O(10^{-4})$



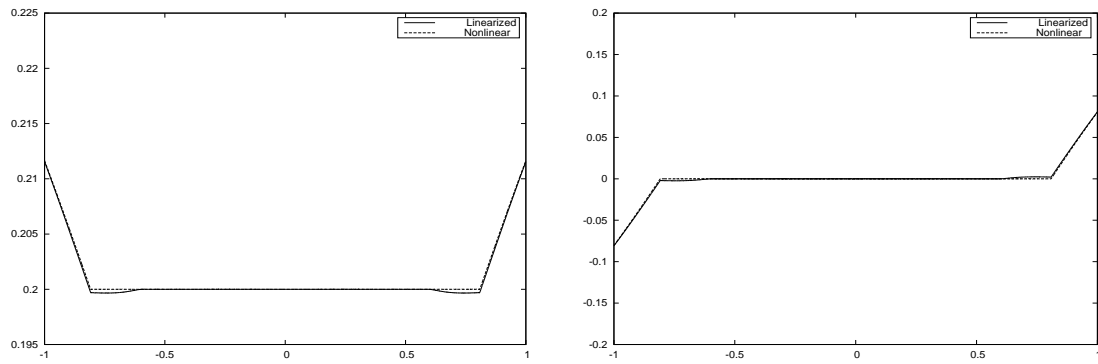
(a) h and u at $t = 0.2$



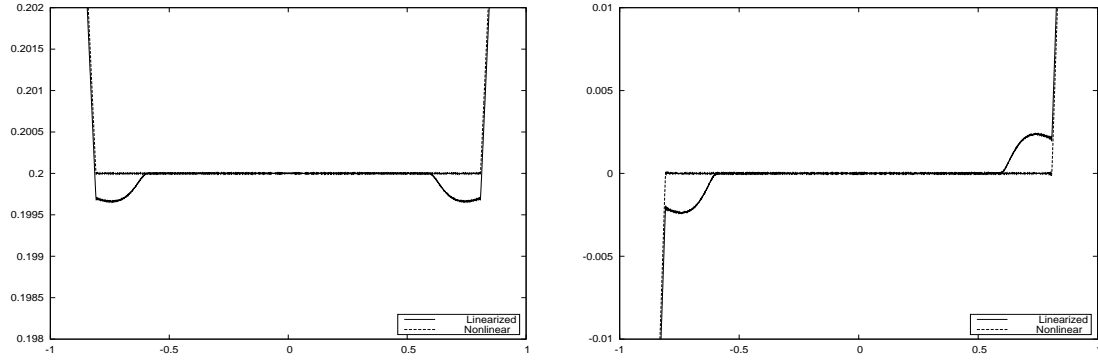
(b) h and u at $t = 0.4$



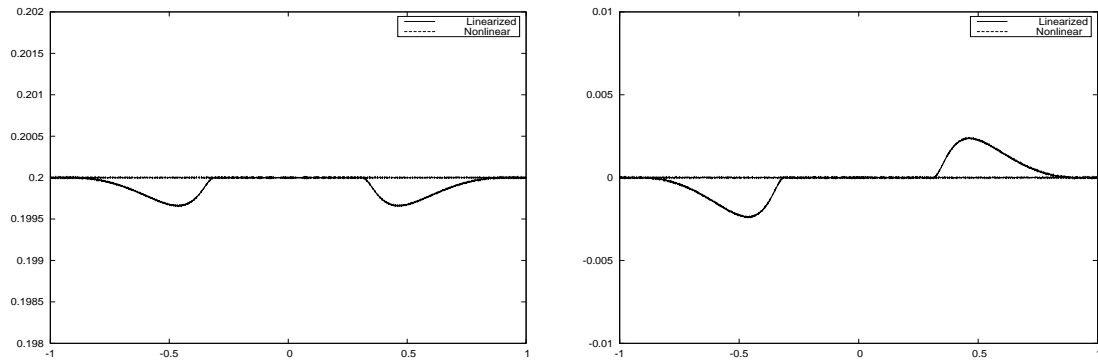
(c) h and u at $t = 0.6$



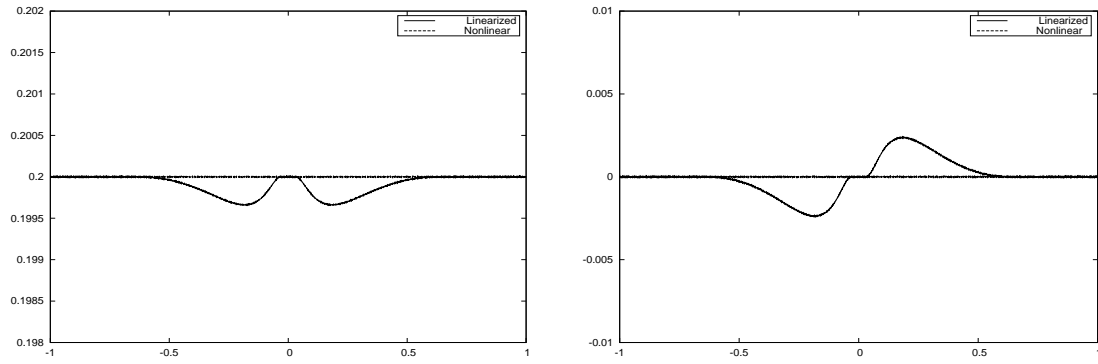
(d) h and u at $t = 0.8$



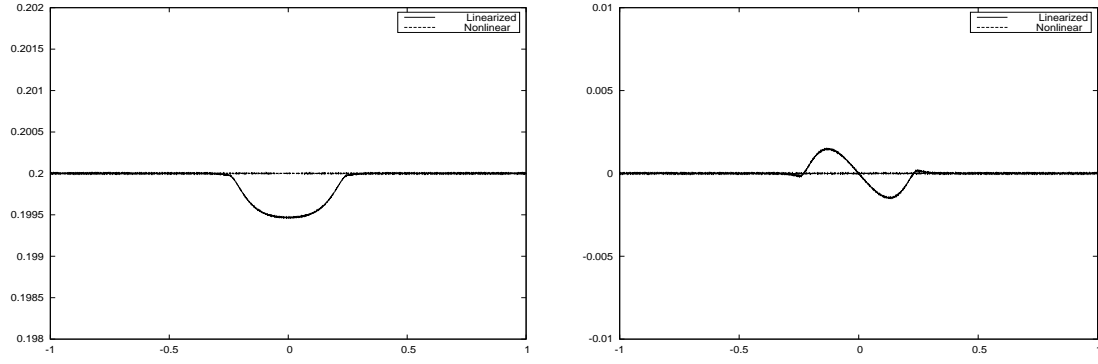
(e) magnification of (d)



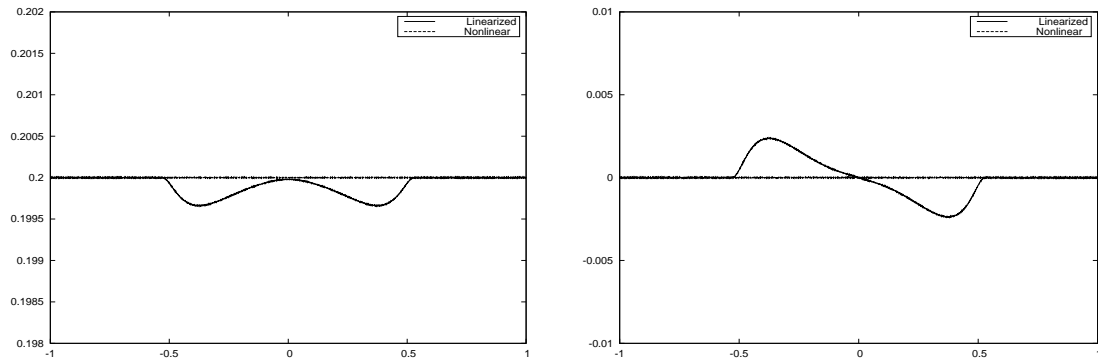
(f) h and u at $t = 1.0$



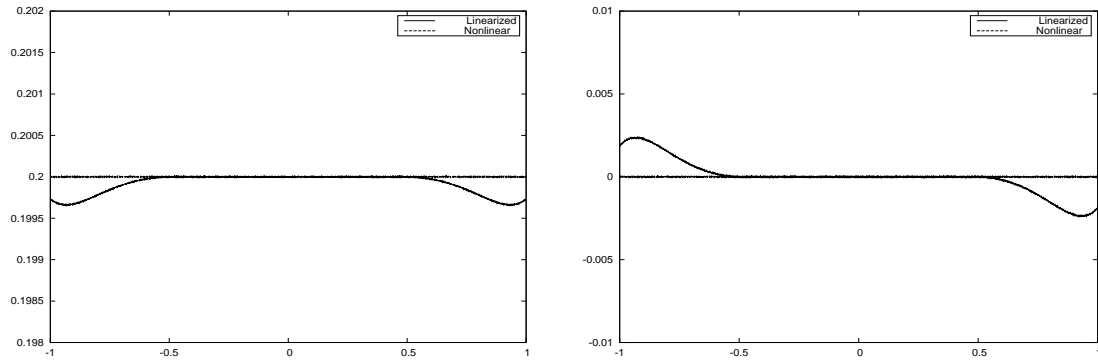
(g) h and u at $t = 1.2$



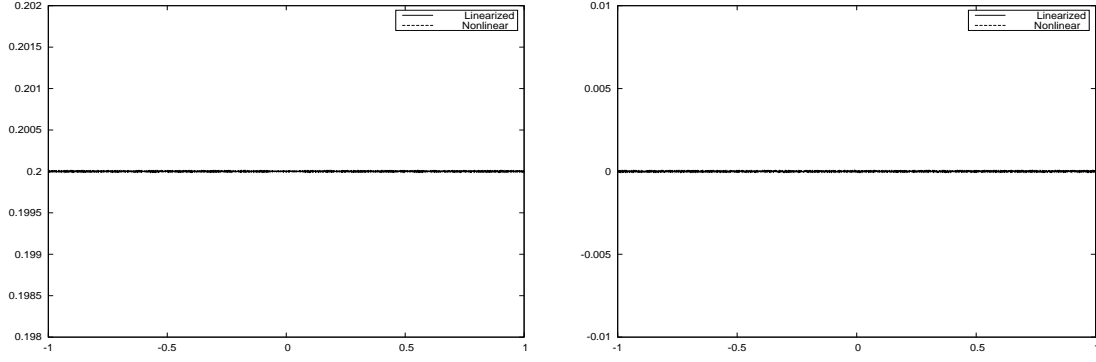
(h) h and u at $t = 1.4$



(i) h and u at $t = 1.6$



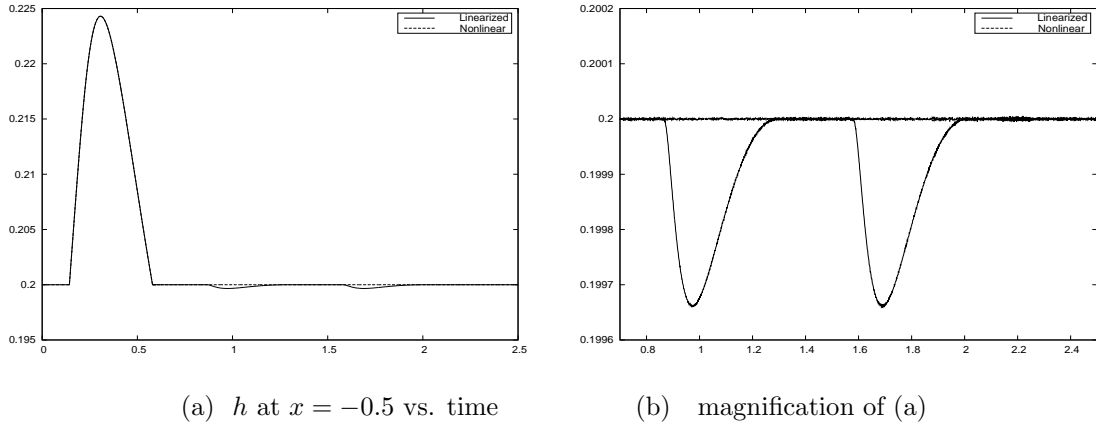
(j) h and u at $t = 2.0$



(k) h and u at $t = 2.4$

FIGURE 4.8. Evolution of $h(x,t)$ (left) and $u(x,t)$ (right); numerical solution of (4.9)-(4.10) on $[-1, 1]$ with nonlinear (dotted line) and linearized (solid line) outflow b.c.'s with $v(x) = 0$, $f(x) = 0.2 + 0.05 \sin(\pi(x + 0.3)/0.6)$ if $|x| \leq 0.3$, $f(x) = 0.2$ if $|x| \geq 0.3$.

that travel and interact approximately linearly until they exit the domain at about $t = 2.4$ sec. The spurious reflections may also be observed in the graphs of the temporal history of the waveheight at a gauge at $x = -0.5$ m computed with the nonlinear (dotted line) and the linearized (solid line) boundary conditions.



(a) h at $x = -0.5$ vs. time

(b) magnification of (a)

FIGURE 4.9. Graphs of $h(-0.5, t)$ for $0 \leq t \leq 2.5$; evolution of Figure 4.8. Nonlinear b.c.'s (dotted line) vs. linearized b.c.'s (solid line).

The passage of the main pulse at the gauge location is completed at about $t = 0.6$ sec. Subsequently the rightwards-travelling reflected spurious small-amplitude pulse due to the linearized b.c.'s is observed to be passing the gauge location at around $t = 1$ sec, while the leftwards-travelling companion spurious pulse passes at around $t = 1.7$ sec.

This experiment is qualitatively analogous to that described in Section 4.2 of [NMF]. We simulated here a smaller-amplitude wave to avoid the emergence of discontinuities during the temporal interval of the evolution of Fig. 4.8. In [NMF] the analogous maximum waveheight was equal to about 0.3 as the finite-difference scheme used in that work had artificial viscosity (private communication by Dr. A. McC. Hogg), that smoothed out the emerging discontinuity. (A more correct simulation of the numerical experiment in [NMF] requires that an initial half-sinusoidal *impulse* is imparted on the water column. This may be modelled by an appropriate forcing term of the form $-F(x,t)/h$ on the right-hand side of the momentum equation, i.e. the second pde in (4.9), where $F(x,t)$ may be taken as a multiple of $\sin(\pi t)$ if $(x,t) \in [-0.1, 0.1] \times [0, 1]$ and $F = 0$ otherwise, while the initial conditions should now be $f(x) = 0.2$ m and $v(x) = 0$. For small enough amplitude of the forcing term F we observed that a qualitatively similar evolution took place.)

Acknowledgement. The authors would like to thank Prof. J. Nycander, Dr. A.McC. Hogg, and Dr. L.M. Frankcombe for helpful comments on their numerical experiments in [NMF].

REFERENCES

- [AD1] D.C. Antonopoulos and V.A. Dougalis, Error estimates for Galerkin approximations of the ‘classical’ Boussinesq system, *Math. Comp.* 82(2013), 689–717. (Extended version in arXiv:100.4248).
- [AD2] D.C. Antonopoulos and V.A. Dougalis, Error estimates for the standard Galerkin-finite element method for the Shallow Water equations (To appear in *Math. Comp.*; extended version in arXiv:1403.5699)
- [BDMK] Bona, J. L., Dougalis, V. A., Karakashian, O. A. & McKinney, W. R., Conservative, high-order numerical schemes for the generalized Korteweg de-Vries equation, *Phil. Trans. R. Soc. Lond. A* 351(1995), 107–164.
- [BPSTT] A. Bousquet, M.Petcu, M.-C. Shiue, R. Temam, and J. Tribbia, Boundary conditions for limited area models based on the Shallow Water equations, *Commun. Comput. Phys.* 14(2013), 664–702.
- [DDW] J. Douglas, Jr., T. Dupont, and L. Wahlbin, Optimal L^∞ error estimates for Galerkin approximations to solutions of two-point boundary value problems, *Math. Comp.* 29(1975), 475–483.
- [D1] T. Dupont, Galerkin methods for first order hyperbolics: an example, *SIAM J. Numer. Anal.* 10(1973), 890–899.
- [D2] T. Dupont, Galerkin methods for modelling gas pipelines, *Lecture Notes in Math.*, v. 430(1974), 112–130, Springer-Verlag, Berlin, 1974.
- [FM] L.M. Frankcombe and A. McC. Hogg, Tidal modulation of two-layer hydraulic exchange flows, *Ocean Sci.* 3(2007), 179–188.
- [HPT] A. Huang, M. Petcu, and R. Temam, The one-dimensional supercritical Shallow Water Equations with topography, *Ann. Univ. Bucharest (Mathematical Series) 2 (LX)* (2011), 63–82.
- [KNP] A. Kurganov, S. Noelle, and G. Petrova, Semidiscrete central-upwind schemes for conservation laws and Hamilton-Jacobi equations, *SIAM J.Sci.Comput.* 23(2001), 707–740.
- [KP] A. Kurganov and G. Petrova, Central schemes and contact discontinuities, *ESAIM-M2AN*, 34(2000), 1259–1275.
- [Ma] A. Majda, *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*, Springer-Verlag, New York 1984.
- [ND] J. Nycander and K. Döös, Open boundary conditions for barotropic waves, *J. Geophys. Res.*, 108(2003), doi:10.1029/2002JC001529.
- [NMF] J. Nycander, A.McC. Hogg, and L.M. Frankcombe, Open boundary conditions for nonlinear channel flow, *Ocean Modelling* 24(2008), 108–121.
- [PT1] M. Petcu and R. Temam, The one-dimensional shallow water equations with transparent boundary conditions, *Math. Meth. Appl. Sci.* 2011, doi:10.1002/mma.1482.
- [PT2] M. Petcu and R. Temam, An interface problem : the two-layer shallow water equations, *DCDS-A* 33(2013), 5327–5345.
- [Sch] R. Schreiber, Finite element methods of high-order accuracy for singular two-point boundary value problems with non-smooth solutions, *SIAM J. Numer. Anal.* 17(1980), 547–566.
- [SLTT] M.-C. Shiue, J. Laminie, R. Temam, and J. Tribbia, Boundary value problems for the shallow water equations with topography, *J. Geophys. Res.* 116, C02015, doi:10.1029/2010JC006315.
- [Wh] G.B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York 1974.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ATHENS, 15784 ZOGRAPHOU, GREECE, AND INSTITUTE OF APPLIED AND COMPUTATIONAL MATHEMATICS, FORTH, 70013 HERAKLION, GREECE
E-mail address: antonod@math.uoa.gr , doug@math.uoa.gr